



# Legacy ExtraGalactic UV Survey with The *Hubble Space Telescope*: Stellar Cluster Catalogs and First Insights Into Cluster Formation and Evolution in NGC 628\*

A. Adamo<sup>1</sup>, J. E. Ryon<sup>2,3</sup>, M. Messa<sup>1</sup>, H. Kim<sup>4</sup>, K. Grasha<sup>5</sup>, D. O. Cook<sup>6,22</sup>, D. Calzetti<sup>5</sup>, J. C. Lee<sup>2,7</sup>, B. C. Whitmore<sup>2</sup>, B. G. Elmegreen<sup>8</sup>, L. Ubeda<sup>2</sup>, L. J. Smith<sup>9</sup>, S. N. Bright<sup>2</sup>, A. Runnholm<sup>1</sup>, J. E. Andrews<sup>10</sup>, M. Fumagalli<sup>11</sup>, D. A. Gouliermis<sup>12,13</sup>, L. Kahre<sup>14</sup>, P. Nair<sup>15</sup>, D. Thilker<sup>16</sup>, R. Walterbos<sup>14</sup>, A. Wofford<sup>17</sup>, A. Aloisi<sup>2</sup>, G. Ashworth<sup>11</sup>, T. M. Brown<sup>2</sup>, R. Chandar<sup>18</sup>, C. Christian<sup>2</sup>, M. Cignoni<sup>19,20</sup>, G. C. Clayton<sup>21</sup>, D. A. Dale<sup>22</sup>, S. E. de Mink<sup>23</sup>, C. Dobbs<sup>24</sup>, D. M. Elmegreen<sup>25</sup>, A. S. Evans<sup>26,27</sup>, J. S. Gallagher III<sup>3</sup>, E. K. Grebel<sup>28</sup>, A. Herrero<sup>29,30</sup>, D. A. Hunter<sup>31</sup>, K. E. Johnson<sup>26</sup>, R. C. Kennicutt<sup>32</sup>, M. R. Krumholz<sup>33</sup>, D. Lennon<sup>34</sup>, K. Levay<sup>2</sup>, C. Martin<sup>6</sup>, A. Nota<sup>9</sup>, G. Östlin<sup>1</sup>, A. Pellerin<sup>35</sup>, J. Prieto<sup>36</sup>, M. W. Regan<sup>2</sup>, E. Sabbi<sup>2</sup>, E. Sacchi<sup>37,38</sup>, D. Schaerer<sup>39</sup>, D. Schiminovich<sup>40</sup>, F. Shabani<sup>28</sup>, M. Tosi<sup>38</sup>, S. D. Van Dyk<sup>38</sup>, and E. Zackrisson<sup>41</sup>

<sup>1</sup> Department of Astronomy, Oskar Klein Centre, Stockholm University, AlbaNova University Centre, SE-106 91 Stockholm, Sweden; [adamo@astro.su.se](mailto:adamo@astro.su.se)

<sup>2</sup> Space Telescope Science Institute, Baltimore, MD, USA

<sup>3</sup> Department of Astronomy, University of Wisconsin–Madison, Madison, WI, USA

<sup>4</sup> Gemini Observatory, La Serena, Chile

<sup>5</sup> Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

<sup>6</sup> California Institute of Technology, Pasadena, CA, USA

<sup>7</sup> Visiting Astronomer, Spitzer Science Center, Caltech, Pasadena, CA, USA

<sup>8</sup> IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY, USA

<sup>9</sup> European Space Agency/Space Telescope Science Institute, Baltimore, MD, USA

<sup>10</sup> Department of Astronomy, University of Arizona, Tucson, AZ, USA

<sup>11</sup> Institute for Computational Cosmology and Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham, UK

<sup>12</sup> Zentrum für Astronomie der Universität Heidelberg, Institut für Theoretische Astrophysik, Albert-Ueberle-Str. 2, D-69120 Heidelberg, Germany

<sup>13</sup> Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany

<sup>14</sup> Department of Astronomy, New Mexico State University, Las Cruces, NM, USA

<sup>15</sup> Department of Physics and Astronomy, University of Alabama, Tuscaloosa, AL, USA

<sup>16</sup> Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD, USA

<sup>17</sup> Instituto de Astronomía, Universidad Nacional Autónoma de México, Unidad Académica en Ensenada, Km 103 Carr. Tijuana-Ensenada, Ensenada 22860, México

<sup>18</sup> Department of Physics and Astronomy, University of Toledo, Toledo, OH, USA

<sup>19</sup> Department of Physics, University of Pisa, Largo B. Pontecorvo 3, I-56127, Pisa, Italy

<sup>20</sup> INFN, Largo B. Pontecorvo 3, I-56127, Pisa, Italy

<sup>21</sup> Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA, USA

<sup>22</sup> Department of Physics and Astronomy, University of Wyoming, Laramie, WY, USA

<sup>23</sup> Astronomical Institute Anton Pannekoek, Amsterdam University, Amsterdam, The Netherlands

<sup>24</sup> School of Physics and Astronomy, University of Exeter, Exeter, UK

<sup>25</sup> Department of Physics and Astronomy, Vassar College, Poughkeepsie, NY, USA

<sup>26</sup> Department of Astronomy, University of Virginia, Charlottesville, VA, USA

<sup>27</sup> National Radio Astronomy Observatory, Charlottesville, VA, USA

<sup>28</sup> Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr. 12-14, D-69120 Heidelberg, Germany

<sup>29</sup> Instituto de Astrofísica de Canarias, La Laguna, Tenerife, Spain

<sup>30</sup> Departamento de Astrofísica, Universidad de La Laguna, Tenerife, Spain

<sup>31</sup> Lowell Observatory, Flagstaff, AZ, USA

<sup>32</sup> Institute of Astronomy, University of Cambridge, Cambridge, UK

<sup>33</sup> Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT Australia

<sup>34</sup> European Space Astronomy Centre, ESA, Villanueva de la Cañada, Madrid, Spain

<sup>35</sup> Department of Physics and Astronomy, State University of New York at Geneseo, Geneseo, NY, USA

<sup>36</sup> Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA

<sup>37</sup> Department of Physics and Astronomy, Bologna University, Bologna, Italy

<sup>38</sup> INAF—Osservatorio Astronomico di Bologna, Bologna, Italy

<sup>39</sup> Observatoire de Genève, University of Geneva, Geneva, Switzerland

<sup>40</sup> Department of Astronomy, Columbia University, New York, NY, USA

<sup>41</sup> Department of Physics and Astronomy, Uppsala University, Box 515, SE-751 20 Uppsala, Sweden

Received 2017 February 14; revised 2017 April 28; accepted 2017 April 30; published 2017 June 5

## Abstract

We report the large effort that is producing comprehensive high-level young star cluster (YSC) catalogs for a significant fraction of galaxies observed with the Legacy ExtraGalactic UV Survey (LEGUS) *Hubble* treasury program. We present the methodology developed to extract cluster positions, verify their genuine nature, produce multiband photometry (from NUV to NIR), and derive their physical properties via spectral energy distribution fitting analyses. We use the nearby spiral galaxy NGC 628 as a test case for demonstrating the impact that LEGUS will have on our understanding of the formation and evolution of YSCs and compact stellar associations within their host galaxy. Our analysis of the cluster luminosity function from the UV to the NIR finds a steepening at the bright end and at all wavelengths suggesting a dearth of luminous clusters. The cluster mass function of NGC 628

\* Based on observations obtained with the NASA/ESA *Hubble Space Telescope*, at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

is consistent with a power-law distribution of slopes  $\sim -2$  and a truncation of a few times  $10^5 M_\odot$ . After their formation, YSCs and compact associations follow different evolutionary paths. YSCs survive for a longer time frame, confirming their being potentially bound systems. Associations disappear on timescales comparable to hierarchically organized star-forming regions, suggesting that they are expanding systems. We find mass-independent cluster disruption in the inner region of NGC 628, while in the outer part of the galaxy there is little or no disruption. We observe faster disruption rates for low mass ( $\leq 10^4 M_\odot$ ) clusters, suggesting that a mass-dependent component is necessary to fully describe the YSC disruption process in NGC 628.

*Key words:* galaxies: individual (NGC 628, M74) – galaxies: star clusters: general – galaxies: star formation – stars: formation

## 1. Introduction

Young star cluster (YSC) populations, commonly detected in local star-forming galaxies, can be powerful tracers of the star formation process in their host galaxies. YSCs form in and interact with their host galaxies, and as bound objects they allow us to study the star formation histories (SFHs) of their parent galaxies (e.g., Adamo et al. 2010a; Glatt et al. 2010). It is intriguing to find very ancient gravitationally bound stellar objects, i.e., the globular clusters (GCs). Potentially, GCs and YSCs could share the same formation process, although the former have most likely formed under different interstellar medium (ISM) physical conditions (e.g., Swinbank et al. 2010). Both similarities and differences between the young and ancient cluster populations are numerous, and it is not clear yet whether YSCs could evolve into GCs (e.g., Kruijssen 2015).

Through the access to multiband *HST* data sets, it has been possible to conduct several studies of YSC populations in local galaxies. YSCs appear to be a common product of star formation in local galaxies. They form in the densest regions of the giant molecular clouds (GMCs), nested at the bottom of the hierarchically structured ISM (Elmegreen & Efremov 1997). Therefore, they can be used to probe star formation processes in local galaxies.

Investigating properties like the mass function (or the luminosity function) of YSC populations can help to constrain their formation mechanism and how they are linked to the overall star formation process in galaxies. Throughout the literature, there is consensus and evidence that both the initial cluster mass and luminosity functions (CMF and CLF, respectively) are well described by a power-law slope of approximately  $-2$  (Whitmore et al. 1999; Larsen 2002; Bik et al. 2003; Gieles et al. 2006; Mora et al. 2009; Whitmore et al. 2010; Chandar et al. 2014; Whitmore et al. 2014, among many others). It has also been observed that the range for the recovered CLF slopes in several nearby galaxies is quite large,  $-2.8 < \delta < -1.5$  (e.g., Adamo et al. 2011; Annibali et al. 2011; Whitmore et al. 2014). Blending effects, important in crowded fields and in galaxies at large distances (above 80 Mpc), have the tendency to flatten the slope of the CLF (Randriamanakoto et al. 2013). However, in some dwarf starburst galaxies (Adamo et al. 2010a, 2011; Annibali et al. 2011), the CLF slopes appear to be flatter than those of rich YSC populations in, e.g., M51 and the Antennae over the same luminosity range. On the other hand, steeper slopes have been observed as a function of wavelength (e.g., Haas et al. 2008), and at brighter magnitude ranges (e.g., Gieles 2010; Bastian et al. 2012) in some local spiral galaxies. This steepening suggests that we find a smaller number of luminous clusters than expected if the luminosity function results from an underlying mass function described by a power law with slope

$-2$  and no upper mass limits. The dearth of very luminous (thus, massive) clusters could be explained if the CMF were not a pure power law, but took the form of a Schechter function, which includes an exponential truncation at masses above  $M_\star$  (e.g., Gieles et al. 2006; Larsen 2009). However, the true shape and universality of the CMF still remains under debate and requires the investigation of a significantly larger sample of galaxies.

Another key aspect not yet fully understood is whether or not there is a change in the cluster formation efficiency ( $\Gamma$ , the mass of star formation that is locked into star clusters) as a function of galactic environment (e.g., Adamo & Bastian 2015). Observational evidence suggests an increase in the cluster formation efficiency as a function of SFR density,  $\Sigma_{\text{SFR}}$  (e.g., Goddard et al. 2010; Adamo et al. 2011, 2015). A model proposed by Kruijssen (2012) suggests that this trend is produced by the link between the cluster formation efficiency and the properties of the hierarchically structured ISM. In this model, YSCs form in regions that reach gas densities above a critical value. Higher gas pressures (thus, higher gas surface densities) will favor the formation of clusters (Adamo et al. 2015). Since gas surface density is directly linked to the  $\Sigma_{\text{SFR}}$  via the Schmidt–Kennicutt (Kennicutt & Evans 2012) relation, it can explain the observed increase of cluster formation efficiency. Evidence has also been reported (Chandar et al. 2015) that  $\Gamma$  does not strongly correlate with total SFR, suggesting that when SFR is used instead of  $\Sigma_{\text{SFR}}$ , the environmental dependency becomes a second-order effect.

The nature of this type of relation is fundamental for understanding the clustering properties of star formation. For example, blue compact galaxies dominated by a recent starburst appear to have a cluster formation efficiency above 40% (Adamo et al. 2011). Since the majority of the massive stars (the dominant source of ionizing photons) is forming in clusters, cluster feedback has a large impact on the ISM of these galaxies (e.g., Bik et al. 2015). The  $\Gamma$  versus  $\Sigma_{\text{SFR}}$  relation can also explain why dwarf irregulars living in galaxy clusters host significantly larger samples of GCs than their counterpart dwarf systems which have spent most of their time in the field. The former, entering the overdense regions, have experienced a starburst event that has produced numerous clusters (Mistani et al. 2016).

A full description of YSC populations in local galaxies also requires a clear understanding of their evolution in the host galaxies. Two main scenarios are currently considered, and analyses based on the YSC populations of the same galaxies have not reached an agreement (see, e.g., Adamo & Bastian 2015 for a summary). The disruption model put forward by, e.g., Fall et al. (2005) is historically based on the disruption rate recovered from the age distributions of YSCs in the Antennae

galaxy (see Whitmore et al. 2010 for the latest analysis). It proposes that YSCs rapidly dissolve (80%–90% each age dex) first because of internal evaporation (e.g., two-body relaxation, stellar evolution) and on long timescales because of external (e.g., tidal fields) processes. Fall et al. (2009) have suggested that these processes happen over different timescales and are independent of the cluster masses. The other scenario is described in Lamers et al. (2005) and observationally supported by recovered disruption rates in the solar neighbors SMC, M33, and M51 (e.g., Boutloukos & Lamers 2003; Gieles et al. 2005). It suggests that YSCs do not disrupt so rapidly, but their mass losses become significant after some time because of interactions with the GMCs, host tidal fields, and stellar evolution (Portegies Zwart et al. 2010). Disruption, in this latter scenario, is dependent on the mass of the YSCs, with low mass clusters disrupting more rapidly. There is another important difference between the two empirical scenarios described above, i.e., the role of the galactic environment. The mass-independent scenario proposes a “quasi-universal” disruption rate independent of the cluster mass and galactic environment where clusters are formed, indicating that the primary disruptive influences may be due to internal processes. The other scenario is clearly anchored to the differences in the properties of the host galaxies. Theoretical models (Elmegreen & Hunter 2010; Kruijssen et al. 2011) suggest that interactions with dense ISM (clusters form in GMC complexes, which will exert tidal forces on the clusters) can also disrupt clusters of any mass. Therefore, we should see that the disruption rate should change as a function of environment (see Adamo & Bastian 2015 for a compilation of observational evidence).

Many fundamental questions related to cluster formation and evolution still remain unanswered. The Legacy ExtraGalactic UV survey (LEGUS) is a Hubble treasury program designed to provide a homogeneous imaging data set in five bands (from the UV to the NIR) of a large sample (50) of local star-forming galaxies, representative of the variety observed within the Local Volume (Calzetti et al. 2015, hereafter C15). The homogeneous imaging coverage, including two filters below the Balmer break ( $\sim 4000$  Å), is enabling us to recover high-quality cluster photometric and physical properties for a large number of YSC populations in local galaxies. In the pre-LEGUS era, only a handful of galaxies can claim comparable data and products, like M31 (Johnson et al. 2015), the Antennae (Whitmore et al. 2010), and M83 (Chandar et al. 2014; Adamo et al. 2015). The access to a large sample of galaxies with high-quality cluster catalogs produced with the most up-to-date techniques and homogeneous approaches will provide a common ground on which to investigate and try to answer all the open questions addressed above.

The aim of this work is to present the LEGUS cluster analysis and provide guidelines to the numerous cluster catalogs that will be released to the astronomical community in 2017. We use as a test bench the nearby spiral galaxy NGC 628 (also known as M74, distance of  $\sim 9.9$  Mpc from C15; see Figure 1). Morphologically, this galaxy has been classified as a Hubble type SAc galaxy<sup>42</sup>, a multiple spiral-arm system. Previous studies of NGC 628 (e.g., Lelièvre & Roy 2000; Thilker et al. 2002) find a HII-region luminosity function slightly shallower than  $-2$  and no significant change of slope as a function of galactocentric distances but only between arm and inter-arm regions (Kennicutt et al. 1989). Elmegreen et al. (2006)

report luminosity and mass distributions of increasing stellar aggregate boundaries compatible with a power-law index of  $-2$ , suggesting that star formation in the NGC 628 inner and outer disk proceeds in a scale-free hierarchical fashion as result of a turbulence-dominated ISM. The LEGUS data set (see Figure 1) now available offers us the possibility to continue this investigation at the smallest and yet densest star-forming scales, at the bottom of this hierarchical process.

In the first part of this paper, we provide a thorough description of the methodology developed to extract cluster candidate positions, produce final photometric tables, and investigate completeness limits of our data set (Section 2). In Section 3, we describe the fitting methods used to derive cluster candidate physical properties. The color properties of the cluster candidates of our test galaxy NGC 628 are analyzed in Section 4. In Section 5, we present the CLF, CMF, and disruption rate analysis. In the final sections, we discuss our results in the framework of cluster formation and evolution and summarize the main result in the conclusions.

## 2. The Photometric Cluster Catalogs

### 2.1. Data Set Description

We refer the reader to C15 for a description of the standard data reduction of the LEGUS imaging data sets, which are currently released at the Web site <https://archive.stsci.edu/prepds/legus/>. Each LEGUS target has a standard and homogeneous filter coverage provided by archival and newly acquired imaging with either the Wide Field Camera 3 (WFC3) or the Advanced Camera for Surveys (ACS). All of the galaxies have WFC3 imaging in the F275W and F336W filters. Three other bands cover the blue optical (ACS/F435W or WFC3/F438W), visual (ACS or WFC3 F555W or F606W), and NIR (ACS or WFC3 F814W). Hereafter, the conventional Johnson passband naming the *UV*, *U*, *B*, *V*, and *I* band will be used for the five filters, although the filter throughput is not converted to the standard Johnson system. We refer hereafter to the *V* band as the reference frame. The photometry provided by the LEGUS analysis is in the Vega magnitude system. Reduced science frames have been drizzled to a common scale resolution, which corresponds to the native WFC3 pixel size ( $0''.04/\text{px}$ ). The frames have all been aligned and rotated north up.

A description of the LEGUS imaging available for NGC 628 is given in Table 1. The data set consists of a mixture of archival and newly acquired data that have been reduced according to the standard LEGUS approach (see C15 for details).

### 2.2. Extraction and Selection of Cluster Candidates

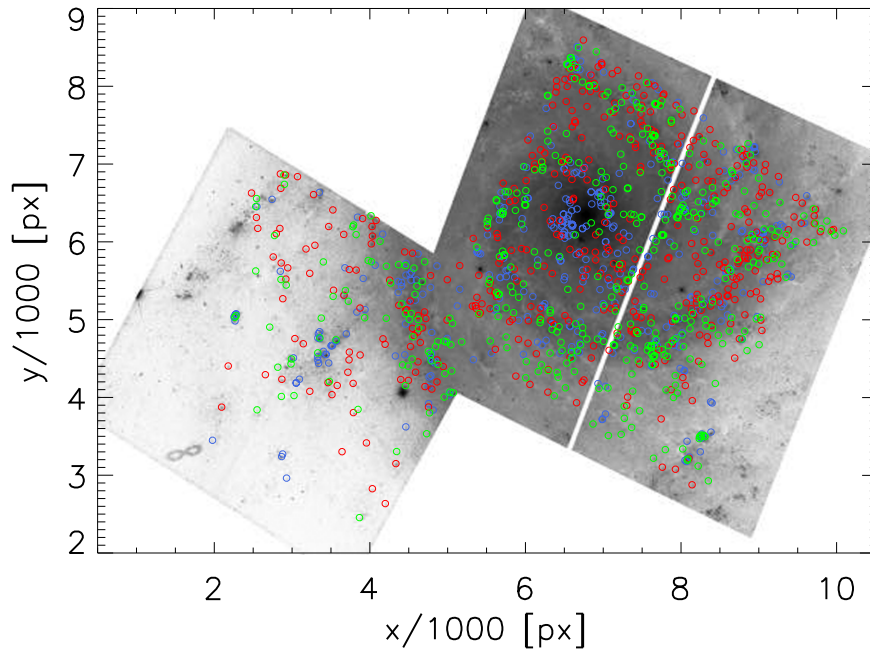
#### 2.2.1. Automated Steps

A custom pipeline, `legus_clusters_extraction.py` (version 4.0), has been developed to produce initial cluster candidate catalogs that contain, for each source, an ID number, the position in pixel coordinates, the final photometry including errors (in the Vega system), the concentration index (CI), and the number of filters in which the source was detected with a photometric error  $\leq 0.3$  mag. A readme file includes all of the key information about the galaxy, the parameters used to build the catalog, and the content of the columns.

The pipeline comprises six consecutive steps and produces several diagnostics that help to fix key parameters to produce the cluster candidate catalogs for each galaxy. We present here a detailed description of each step in the pipeline and its application

<sup>42</sup> Value taken from NED.





**Figure 1.** Mosaic image of the two F555W pointings of NGC 628, covering the inner part of the galaxy and a portion of the outer disk in the southeast (image rotated with north up). The circles show the position of class 1 (red), class 2 (green), and class 3 (blue) cluster candidates. See Section 2.2.2 for a description of our classification used here. Detected objects are covering the portions of the field of view that are in common among the imaging taken in the five standard LEGUS filters.

to NGC 628. This galaxy has been observed in two different pointings, NGC 628c (inner region) and NGC 628e (outer region). The two data sets have been analyzed separately, because of different combinations of cameras and filters (see Table 1).

1. Step 1—Source extraction. The pipeline uses SExtractor (Bertin & Arnouts 1996) to extract source positions from the white-light image produced with the five standard LEGUS bands (see C15 for the method used to produce white-light images). In cases where the white-light image is not optimal (e.g., it shows numerous artifacts at the edges of the single frames, or the *I* band in low stellar density fields is dominated by red giants), we replace it with the reference V-band frame. The SExtractor input parameters are optimized to extract objects with at least a  $3\sigma$  detection in a minimum of five contiguous pixels. These numbers change when the white-light or the V-band filter is used for detection. We do not apply any filtering to the image but we use small background cells to make sure that the strong gradient in the background between the arms and inter-arm regions does not affect our capability of detecting sources. These initial catalogs are visually screened to understand whether potential clusters have been missed and any possible improvement in the SExtractor configuration file can be made. The configuration file used for the source extraction will be released together with the final LEGUS cluster catalogs for each galaxy. At the end of this first step, we detect 6272 (NGC 628c) and 4539 (NGC 628e) candidate clusters.
2. Step 2—Determination of CI and aperture radius parameters. In this step, the user selects a training sample of objects that is clearly stars (i.e., isolated, bright, and with steep luminosity profiles) and another sample that is clearly clusters (i.e., isolated, relatively bright, and with shallow luminosity profiles) using the reference frame (V band). The pipeline performs aperture photometry on these sources

using radii of increasing size (from 1 px to 20 px with a step of 1 px). The local sky background annulus is set at 21 px and is 1 px wide. From this photometry, the pipeline calculates the CI (the magnitude difference between apertures of radius 1 pix and 3 pix) and the curves of growth for the input lists of stars and clusters. In Figure 2, we show an example of the plots produced by the pipeline in Steps 2 and 3. The top-left panel shows the CI distribution recovered for the star and cluster control sample. From this plot, the user selects a CI value that separates the visually selected stars and clusters as cleanly as possible. In the case of NGC 628c, the value we apply is 1.4. This value is an initial guess that can be iteratively adjusted in Step 3. As we will discuss and further investigate in Step 3 and in the completeness analysis, a selection criterion based on the CI cut allows us to remove a significant amount of stars from the catalog, but it will also remove compact clusters that appear unresolved on the frame (see Section 2.3 for a detailed discussion). The middle panel of Figure 2 shows the growth curve analysis on the stellar and cluster reference sample. The median stellar growth curve (thick red line in the plot) shows how the flux increases for an averaged unresolved source on this specific reference frame. This curve is linked to the stellar PSF produced by a given combination of detector and filter, and will change from galaxy to galaxy. The median cluster growth curve shows (thick blue line in the plot) how the flux is distributed in a resolved cluster spread function (CSF).<sup>43</sup> The increase in flux is slower in the resolved cluster than in the averaged stellar one, explaining why the

<sup>43</sup> Since the PSF is the spread function of a point-like source, we will refer to the spread function of clusters as the cluster spread function. Notice that a CSF does not mean a cluster resolved in their stellar content. Clusters in the LEGUS galaxies cannot be resolved in their single stars. This is also true for the closest targets, where an unresolved cluster core is typically surrounded by partially resolved single stars.

**Table 1**  
The LEGUS Data Set of NGC 628

Filters	Program Number	PI	exptime (s)	ZP(Vega) (mag)	Aver apcor <sup>a</sup> (mag)	Det Limits <sup>b</sup> (mag)	det threshold (electron s <sup>-1</sup> )
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inner pointing (NGC 628c)							
WFC3/F275W	13364	Calzetti	2481.0	22.632	$-0.817 \pm 0.066$	23.29	0.009
WFC3/F336W	13364	Calzetti	2361.0	23.484	$-0.750 \pm 0.060$	23.91	0.010
ACS/F435W	10402	Chandar	1358.0	25.784	$-0.656 \pm 0.034$	24.93	0.013
ACS/F555W	10402	Chandar	858.0	25.731	$-0.634 \pm 0.034$	25.05	0.021
ACS/F814W	10402	Chandar	922.0	25.530	$-0.751 \pm 0.037$	24.27	0.030
Outer pointing (NGC 628e)							
WFC3/F275W	13364	Calzetti	2361.0	22.632	$-0.795 \pm 0.097$	23.38	0.009
WFC3/F336W	13364	Calzetti	1119.0	23.484	$-0.706 \pm 0.059$	23.48	0.018
ACS/F435W	10402	Chandar	4720.0	25.784	$-0.695 \pm 0.039$	25.26	0.010
WFC3/F555W	13364	Calzetti	965.0	25.816	$-0.740 \pm 0.038$	25.22	0.024
ACS/F814W	10402	Chandar	1560.0	25.530	$-0.843 \pm 0.050$	24.42	0.029

**Notes.**

<sup>a</sup> Averaged aperture corrections used to produce the final AV\_APCOR cluster catalogs.

<sup>b</sup> The listed values correspond to the 90% completeness limits at the detection thresholds listed in column 8. Completeness limits have been estimated using synthetic clusters with sizes larger than 1 pc. See details about the meaning of the recovered completeness values in the main text.

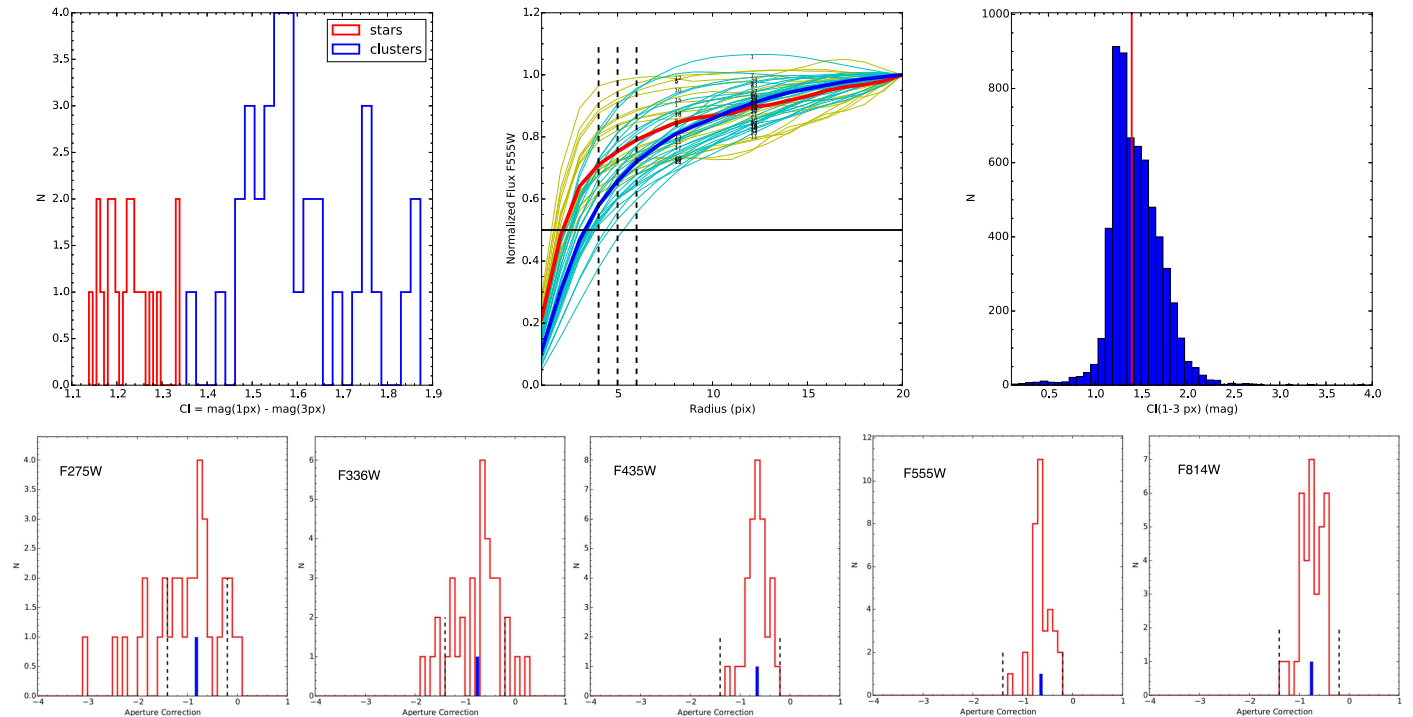
CI is larger for resolved (clusters) than unresolved (stellar-like) sources (the CI method was first introduced by Holtzman et al. 1992; Whitmore et al. 1993). Inspection of the growth curves allows the user to fix the size of the aperture radius used to perform photometry. For each galaxy, we fix the aperture radius to do cluster photometry to the smallest integer number of pixels containing at least 50% of the flux within the aperture (aperture corrections for the missing flux are discussed in Step 4). We select the smallest aperture to reduce as much as possible the risk of contaminations from neighboring sources. We will refer to this value as the science aperture, to distinguish it from other apertures used in the analysis.

- Step 3—CI selection and multiband photometry. The CI distribution of all the extracted sources in Step 1 is plotted together with the CI cut value selected in Step 2. The top-right plot of Figure 2 shows the distribution for NGC 628c. As already discussed in the literature (e.g., Chandar et al. 2010; Adamo et al. 2015), the CI distributions of point sources that are unresolved (stars) will show a narrow peaked distribution (e.g., peak in the right panel of Figure 2). On the other hand, clusters are resolved and they show a broad range of sizes (e.g., Ryon et al. 2015), thus their CI distribution appears significantly broader than the stellar one (this can be seen in the distribution as the prominent wing extended to large CI values). The goodness of the chosen CI value for the cut is then directly tested on the reference frame (V band) via visual inspection of the objects in the catalog that satisfy this condition. We check whether a smaller CI introduces a large contamination of stars. The chosen best CI value is, eventually, indicated as a selection criterion. All sources with a CI smaller than the reference value (1.4 for the ACS V band of NGC 628c and 1.3 for the WFC3 V band of NGC 628e) are removed from the initial catalog. We then perform multiband aperture photometry on the

sources that pass this selection step, using the science aperture fixed in Step 2 and a local sky annulus located at 7 px with a width of 1 px. The same science aperture and local sky annulus is used in all the five filters.

- Step 4—Averaged and CI-based aperture correction. We apply two different aperture correction methods both widely used in the literature.

In the first case, we produce averaged aperture corrections using the cluster control sample produced in Step 2. The correction is estimated as the difference between the magnitude of the source recovered at 20 px (sky annulus at 21 px, 1 px wide) minus the magnitude of the source obtained using the science aperture (sky annulus at 7 px, 1 px wide). Not all of the sources in the control samples are detected in all the filters. To avoid clusters not detected in some filters from skewing the averaged values of the aperture corrections, we reject sources that have corrections outside a certain range. We show for example the aperture correction distributions recovered for the cluster control sample of NGC 628c in the five LEGUS band in the bottom panels of Figure 2. The vertical dashed lines show the limits within which we use the values to produce the average correction. These limits can be adjusted by the user for a given galaxy. In the case of NGC 628c, some of the sources are very faint or not detected in the bluest filters; therefore, their corrections become extreme. In Table 1, we list the final averaged aperture corrections and the errors (standard deviations) for each filter of each pointing in NGC 628. These values are added to the photometry produced in Step 3. The standard deviation of the aperture correction recovered from the sample is added in quadrature to the photometric error to take into account the uncertainties introduced by the averaged correction. With this method, the differences in the sizes of the clusters are not taken into account. However, as one can see from the recovered values in Table 1, the differences in the corrections as a function of



**Figure 2.** Plots produced by the `legus_clusters_extraction.py` pipeline and used to decide important photometric parameters, like CI cut, aperture radius, and range of allowed aperture corrections for each galaxy. The plots included here are for the NGC 628c pointing. Top row: on the left, the recovered CI distributions of the star and cluster control samples are shown. In the middle, the growth curves (normalized flux in  $V$  band vs. aperture) of stars (dark yellow solid lines) and clusters (cyan solid lines) contained in the control samples are plotted. Median curves of both samples are overplotted with red (stars) and blue (clusters) solid lines. The black horizontal solid line shows where the 50% flux is reached. The vertical dashed lines show which fraction of flux is contained in apertures of 4, 5, and 6 px. IDs are included for each single curve, and are used to remove sources that show extreme deviations from the general trends. In the right, the distribution of CI of all the sources extracted in Step 1 is shown. The red vertical line shows the CI reference value used to distinguish between unresolved sources (stars) and resolved objects (candidate clusters). Bottom row: the aperture correction recovered using the cluster control sample. The blue line shows the average value (see Table 1).

waveband are quite small. Therefore, averaged aperture corrections do not change the shape of the spectral energy distribution (SED) of the source—they only change the normalization. This will slightly affect the mass estimates (not the derived ages and extinction values), but it will be well within the average 0.1 dex uncertainties that the SED fitting methods produce.

In the second approach, we produce aperture corrections based on the CI of the source in each band. The method to derive the CI versus aperture correction relation will be described in a forthcoming paper (D. O. Cook et al. 2017, in preparation). This method has initially been developed to produce an aperture correction for very extended clusters (partially resolved in their stellar components) detected in the LEGUS dwarf galaxies and afterwards extended to the rest of the LEGUS targets. The CI-based aperture correction for each source is calculated by measuring the CI in each band and applying the CI–aperture correction relation determined for the appropriate instrument and science aperture size. While this method takes into account the size (indirectly measured via the CI) of the source as a function of waveband, it can change not only the normalization of the SED but also the shape (so it will affect the mass, age, and extinction). This effect will be particularly large in faint sources and in crowded regions where the estimate of the CI becomes more uncertain.

At the end of Step 4, we have two sets of aperture corrections calculated. Within the LEGUS naming convention, the two final photometric catalogs produced in Step 5

with these two correction methods and derived analyses are referred to as `AV_APCOR` and `CI_BASED`.

5. Step 5—Final products. The quality of the photometry is checked in this final step. We remove all of the sources that have not been detected in at least two filters (the reference  $V$  band and either the  $B$  or  $I$  band) with a photometric error  $\sigma_\lambda \leq 0.3$  mag. In each band, we correct the photometry of these sources by the foreground Galactic extinction (Schlafly & Finkbeiner 2011). Two automatic final photometric catalogs are produced, e.g., the `automatic_catalog_ngc628c_avgapcor.tab` and the `automatic_catalog_ngc628c_cibased.tab`. Both catalogs contain the ID of the sources, positions in pixels and in R.A. and decl., multiband photometry and errors, CI, and a flag indicating the number of filters in which each source was detected. In the case of NGC 628c and e, the automatic catalogs produced at the end of this step contain 3086 and 593 cluster candidates, respectively.

These sources, however, are likely not only star clusters. Among them there are still interlopers, i.e., pairs and multiple stars in crowded regions, background galaxies, bright stars in the galaxy that have CI slightly larger than the limit used here, foreground stars, and objects and artifacts at the edge of the science frames. To minimize the contaminations of our final catalogs, we visually inspect a fraction of the sources found in the automatic catalogs that satisfy some extra selection criteria. We select for visual inspection all the sources that are detected in at least four bands (in the `automatic_catalog_ngc628c_avgapcor.tab`)

with a photometric error below 0.3 mag and have an absolute *V*-band magnitude brighter than  $-6$  mag. The *V*-band magnitude cut is applied to the *V*-band photometry obtained with the CI-based aperture correction. This allows us to include slightly fainter, very diffuse clusters since they have large values of the CI and therefore get a larger aperture correction.

6. Step 6—Missed clusters. During the visual classification of the sources it is also possible to add sources that have not been included in the visual inspection catalog. If such extra sources are not found in the automatic catalogs then we use Step 6 to produce their final photometry in the same homogeneous way as done for the bulk of the other sources. The number of added sources is typically below 1% but it can change from galaxy to galaxy.

### 2.2.2. Classification of Cluster Candidates

At this stage of the reduction, three or more independent classifiers from within the LEGUS team visually inspect the cluster candidates. A tool has been made available within the LEGUS collaboration to perform this task. This tool visualizes the object to be classified in two frames, in the reference *V* band and in a combined three-color image. Using an interactive window, the user can also visualize the contour, the radial profile, and the surface plot of each source in the reference band (see Figure 3) using the IRAF task IMEXAMINE. Two circles on the images show the aperture used to perform photometry (inner) and the location of the local sky (outer circle). Each source is inspected in the single-band image to check whether the light is extended (cluster) or has a stellar PSF (star). The difference in the light distribution of a cluster and a star at the distance of NGC 628 can be easily discerned in the top row panels of Figure 3. The source at the center of the inset is a cluster and it has an FWHM = 3.71 px. A star is visible in the same cutout and its light is much less extended (FWHM = 2.4 px) than the one of the cluster. The contour and surface plots (middle and right panels) confirm the different behavior in the two objects.

Based on this inspection, each object in the catalog gets assigned one of four defined classes. In Figure 3 we show an example object for each class taken from the central pointing of NGC 628. Class 1 clusters are compact and centrally concentrated objects, with an FWHM more extended than the stellar one. They show a homogeneous self-consistent color. With respect to class 1, class 2 systems are clusters with slightly elongated density profiles and less symmetric light distribution. Class 3 can potentially be less compact and homogeneous clusters, more likely compact associations<sup>44</sup>, and they show asymmetric profiles and multiple peaks on top of diffuse underlying wings, which suggest the presence of a possible concentration of low mass stars. Finally, class 4 objects are single stars or artifacts, or any other interlopers (e.g., background galaxies, foreground stars) that can affect the cluster analysis. The bottom row of Figure 3 shows an example of an object detected with extended CI, morphology, and CSF profile similar to a class 3 object. However, the stars belonging to this system have clearly different colors. We consider these

objects as a chance superposition along the line of sight of two stars; thus, they are excluded from the cluster catalog.

Within the LEGUS collaboration, we are also testing a machine-learning (ML) recognition of the sources. The visually inspected catalogs are fed to the algorithm as training sets for the subsequent classification of sources. Testing is in progress and the results will be published in a forthcoming paper (Grasha et al. 2017, in preparation). The cluster catalogs of a third of the LEGUS galaxies have currently been inspected by humans. We are providing visual classification of a fraction of the catalogs of another third of galaxies, which are used as training sets by the ML algorithm. The goal is to perform the visual classification of the remaining galaxies using a purely ML-based approach.

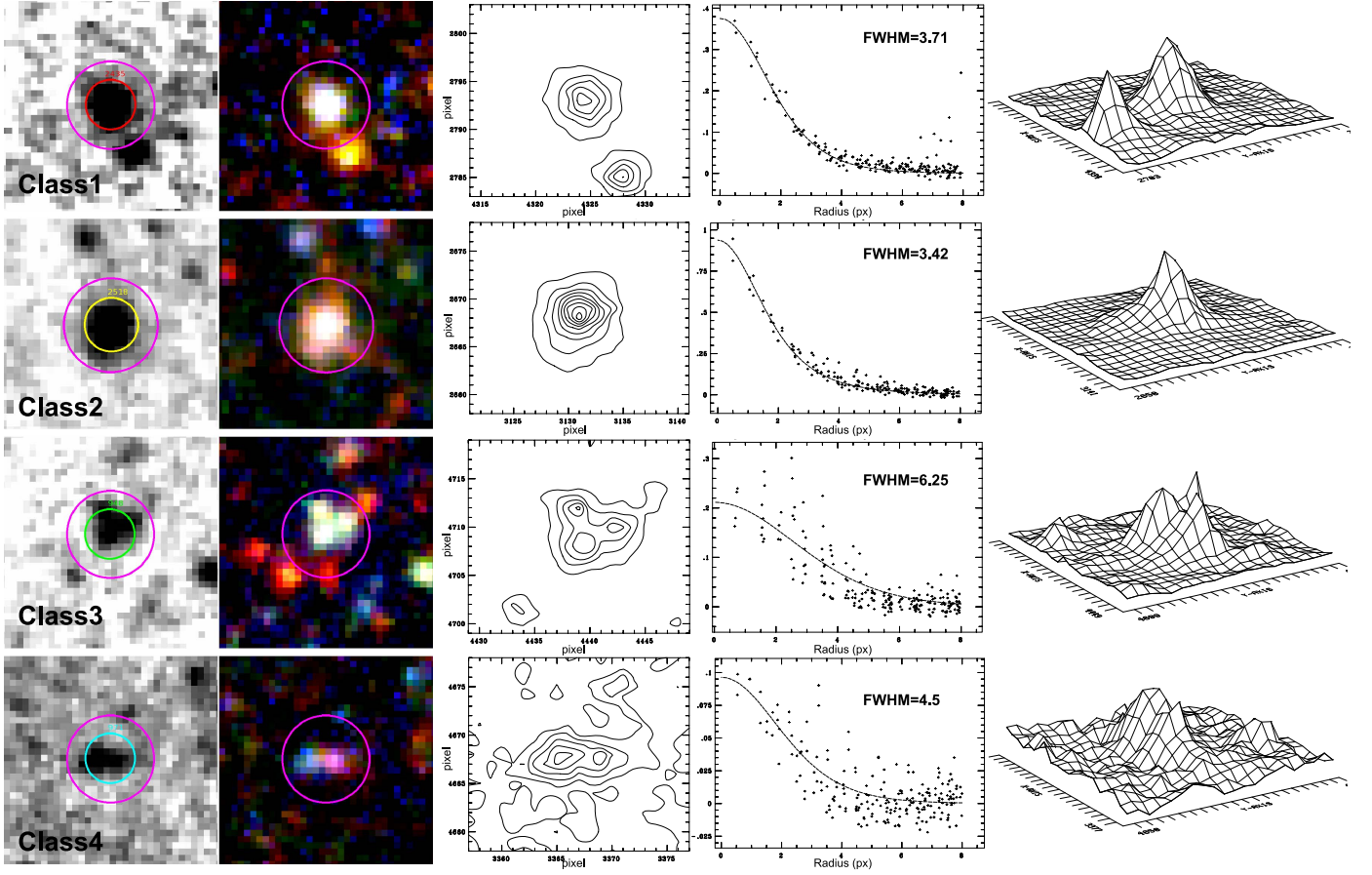
### 2.3. Testing Our Approach and Completeness Limits

To investigate the impact of the assumptions and the selection criteria adopted to produce the LEGUS cluster catalogs, we have performed several completeness tests. The tests are run with a custom-made pipeline, `legus_cct.py` (LEGUS Cluster Completeness Tool), available within the LEGUS collaboration. The pipeline runs in five consecutive steps, i.e., creation of synthetic sources, source extraction, photometry, recovery fractions, and final outputs. One of the input files is the same used to run the standard `legus_clusters_extraction.py`. The second input file specifies quantities necessary to create the synthetic frames containing the simulated clusters, e.g., input parameters for the BAOlab software (Larsen 1999). A stellar PSF generated with the IRAF task PSF using selected stars in the science frames is also provided. This PSF is then convolved with MOFFAT15 profiles of specified effective radii,  $R_{\text{eff}}$ . The resulting extended sources are then randomly distributed in a blank frame of the same dimension as the science one. A magnitude range is provided as well. To overcome crowding problems, we only insert a thousand clusters per loop. The frame containing synthetic clusters is then added to the real science frame. In the next step, all the sources are then extracted using the same procedure as in `legus_clusters_extraction.py`. For all the sources extracted, it estimates the CI and the photometry (including averaged aperture correction) in the same way as for the real clusters. Using the known position of the simulated clusters, the software estimates a source recovery fraction before and after the CI cut is applied.

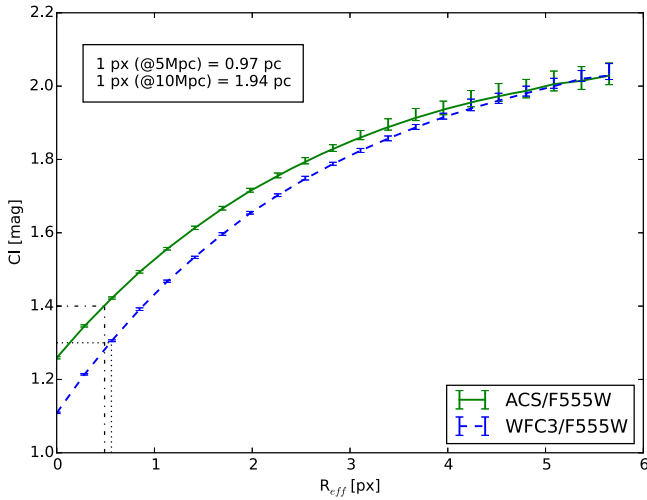
As a first result of these simulations, we have investigated the relation between the CI and  $R_{\text{eff}}$  of increasingly extended sources, with a particular emphasis on the impact that the CI cut has on removing compact clusters as a function of distance of the galaxy. In Figure 4, we show the recovered relation between CI and  $R_{\text{eff}}$  for the ACS and WFC3 F555W science frames, used as reference frames for the CI selection. As an example, we refer to the study case NGC 628. The applied CI cuts of 1.4 mag and 1.3 mag for the ACS and WFC3 F555W correspond to a cluster  $R_{\text{eff}} \sim 1$  pc. This does not imply that all clusters of 1 pc size are systematically removed from the catalogs. We investigate the recovery fraction of sources in the *V*-band frame before and after the CI cut is applied. In Figure 5, we illustrate the results for NGC 628c. Before the CI cut is applied, we have 100% recovery for sources of 1 pc as well as more extended ones. However, the recovery of very compact sources goes below 50% after the CI cut is applied. Only half of the sources with  $R_{\text{eff}} \sim 1$  pc are extended enough to make it

<sup>44</sup> Typical size of stellar associations is between 50 pc and a few hundreds parsecs. Class 3 objects that we will hereafter call associations, or compact associations, have a projected size of a few tens of parsecs and, in many cases, these compact associations are part of much larger stellar associations.

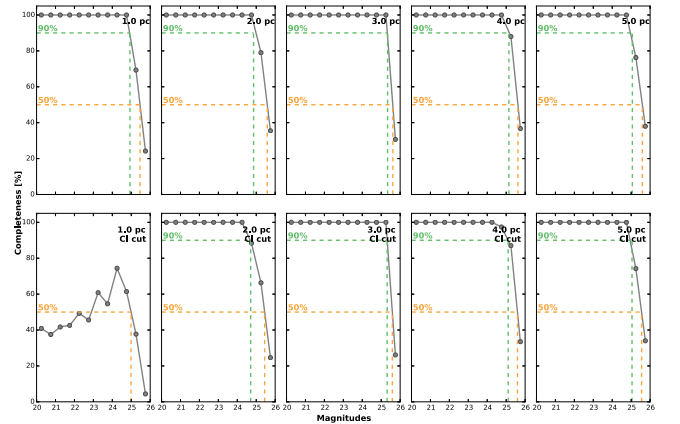




**Figure 3.** Panels inspected for the visual classification of the sources. An example for each identified class is listed. The first two panels of each row show each object in the reference frame *V* band (logarithmic scale) and in a three-color composition. The outer purple ring has a radius of 7 px ( $=0''.28 \sim 13.4$  pc) and shows the position of the local sky annulus. The inner ring shows the aperture radius used to do photometry (4 px  $=0''.16 \sim 7.7$  pc). The middle and right plots show the contour, the radial profile, and the surface plot of the object. The FWHM reported in the radial profile is estimated using a MOFFAT light distribution. Detailed comments on each object are provided in the text.



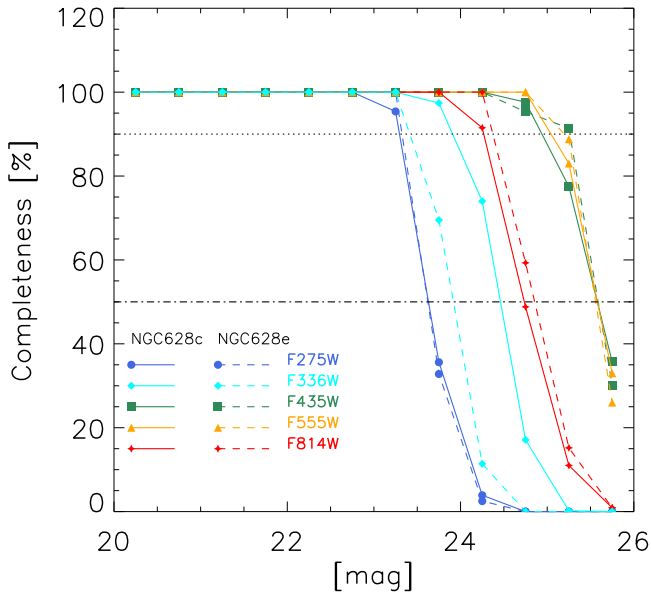
**Figure 4.** Relation between CI and  $R_{\text{eff}}$ . The two curves are the median values obtained from artificial clusters simulated with increasing  $R_{\text{eff}}$ . The relation has been tested for the reference *V* band frame in both the ACS and WFC3 instruments. The error bars are the 25 and 75% quartiles of the measurement. An inset shows the pixel resolution corresponding to the distance range typically covered by the LEGUS galaxies. The black dotted lines show the CI values used for the cluster candidate selection in the inner (ACS/F555W) and outer (WFC3/F555W) of NGC 628. The applied cuts approximately correspond to  $R_{\text{eff}} \sim 1$  pc.



**Figure 5.** Recovered completeness limits in the *V*-band frame of NGC 628c as a function of  $R_{\text{eff}}$  (annotated in each panel). The top row shows the recovery rates before, the bottom row after the CI cut is applied. The CI cut is only affecting the recovery of very compact and partially unresolved clusters with  $R_{\text{eff}} \leq 1$  pc.

into the selection. A smaller CI will include too many stellar objects, so we conclude that at the distance of NGC 628, the CI cut applied removes a fraction of unresolved clusters with sizes of 1 pc and below. This choice does not introduce biases in our cluster analysis because observationally, no clear relation





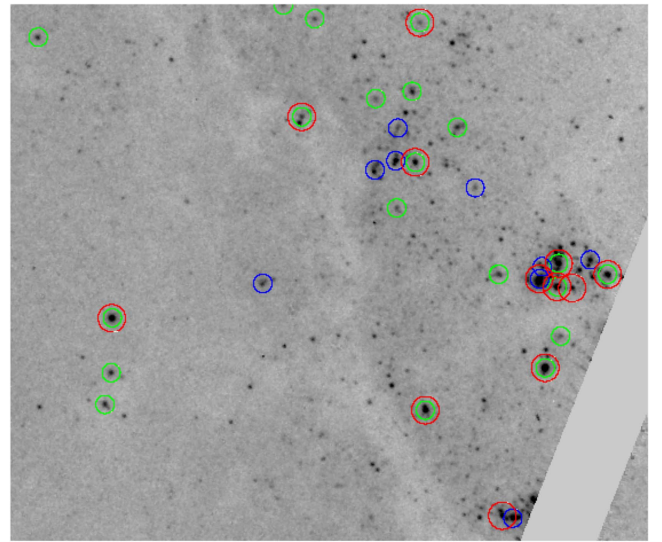
**Figure 6.** Recovered completeness limits as a function of different bands for clusters with sizes larger than 1 pc. Solid and dashed lines associated with different symbols have been used to show the recovery rates in the inner and outer field, respectively (see the inset).

between the size and mass or luminosity of star clusters has been found (see Ryon et al. 2015 and references therein). Moreover, an ongoing analysis of cluster sizes in NGC 628 shows a clear log-normal distribution which is well above the detection limits of 1 pc (Ryon et al. 2017).

In Figure 6, we illustrate the final completeness limits of the two pointings in NGC 628. In Table 1, we list the magnitude limits corresponding to 90% recovery of sources with a detection threshold above  $3\sigma$  (column 8 Table 1) within a minimum of five contiguous pixels. For the same band, the differences in the exposure times (longer exposures) have a larger impact on the recovery fractions than the differences in crowding between the inner and outer region (i.e., compare the recovery magnitude for the F336W and F435W for the inner and outer pointing). The completeness software has been run independently for each band. Therefore, these values must be considered as indicative of the detection limits intrinsic to the science frames. However, we are not taking into account that we apply to our catalogs additional selection constraints (detection in at least two filters, or in the case of visually inspected sources in four filters), visual inspection, and that real clusters have different luminosities at each wavelength depending on their age, mass, and extinction.

#### 2.4. Comparison with Catalogs Available in the Literature

Figure 7 shows a comparison between the LEGUS final catalog (green circles are class 1 and 2, blue circles class 3) and a catalog from Whitmore et al. (2014, shown as red circles) for a small region in NGC 628. The Whitmore et al. catalog was produced automatically based on Hubble Legacy Archive (HLA) observations, including measurements of the concentration index and algorithms that remove close pairs and multiple hits in very crowded regions. A relatively bright limit is used ( $M_V = -8$ ) for the HLA-based catalog since no manual inspections are made.



**Figure 7.** Zoom-in image of a region of NGC 628 ( $560 \times 473$  px). The LEGUS final catalog of objects classified as class 1 and 2 is illustrated by the green circles, class 3 by blue circles. We compare the LEGUS catalog to the automatic catalog (red circle sources) based on HLA observations as described in Whitmore et al. (2014). See the text for differences and similarities in the selection criteria.

The correspondence is fairly good, with 77% of the HLA-based sources being included in the LEGUS catalog. Making a similar magnitude cut in the LEGUS catalog (which goes roughly 2 mag deeper) results in a recovery rate of 73% of the HLA-based catalog being found in the bright end of the LEGUS catalog. There are two primary differences: (1) the LEGUS catalog includes approximately five more (out of about 100 in the region of overlap used for the comparison) bright class 1 objects that are clearly good cluster candidates and were missed in the HLA (probably due to using too conservative a limit for the concentration index), and the LEGUS catalog includes about 15 more compact associations—i.e., type 3—than found in the HLA-based catalog. Although selection effects are an important consideration in any study, we note that Chandar et al. (2014) compared completely manual, hybrid (like with LEGUS), and completely automatic catalogs made by two different studies and found that the selection did not result in major changes to the primary results of the studies (i.e., the CLF, CMF, and age distributions).

### 3. Final Cluster Catalogs and SED Fitting Procedures

The photometric catalogs produced in Step 5 of the pipeline, as described in Section 2.2.1, are fed into two different SED fitting algorithms.

The analysis is performed with two different approaches that reflect the most common methods used in the literature.

In the first case, we fit the cluster SEDs with Yggdrasil SSP models (Zackrisson et al. 2011). We use two stellar libraries to create two sets of SSP models, Padova-AGB and Geneva tracks without rotation, available in Starburst99 (Leitherer et al. 1999; Vázquez & Leitherer 2005). They assume a Kroupa (2001) universal initial mass function (IMF), with stellar masses between 0.1 and  $100 M_\odot$ . The IMF is assumed to be fully sampled; therefore, we will refer to these models as deterministic. These stellar models are then used as input to run Cloudy (Ferland et al. 2013) and obtain a realistic evolution of the nebular emission line and continuum, produced by the

ionized leftover gas surrounding the very young clusters. All the metallicity steps available in Starburst99 for both stellar libraries are accessible. For the Cloudy simulations, we assume that the gas and the stars form in material with the same metallicity. For each galaxy, we adopted the present-day metallicity of its young populations as derived from nebular abundances in the literature and listed in C15. We use spherical solutions for the gas distribution around the ionizing sources. The hydrogen number density,  $n_{\text{H}}$ , is set to typical values measured in H II regions,  $n_{\text{H}} \sim 10^2 \text{ cm}^{-3}$ . The covering factor,  $c$ , is assumed to be 0.5, i.e., roughly 50% of Lyman continuum photons, produced by the central source, ionize the surrounding gas. The gas filling factor,  $f$ , is assumed to be 0.01. While the Yggdrasil interface is able to provide several combinations of Cloudy assumptions, we decided to fix the parameters to average values. Our data set is limited to wide passbands; thus, the total integrated fluxes of these very young clusters are sensitive to the presence of emission lines and continuum from the ionized gas. The changes produced by different assumptions in the gas phases are secondary and difficult to disentangle. Including the nebular treatment in the models is fundamental to estimate the cluster physical parameters of very young stellar clusters (e.g., Zackrisson et al. 2001; Adamo et al. 2010b; Reines et al. 2010); however, we do not have enough information to disentangle the gas conditions. The model grid used in the fitting procedure includes a combination of progressive age steps and increasing internal reddening, i.e.,  $E(B - V) = [0.0; 1.5]$  and steps of 0.01 mag. The models are reddened prior to being fitted to the observed photometry. We provide to the fit three extinction and/or attenuation laws. First, the Milky Way extinction law from Cardelli et al. (1989). Second, the starburst extinction law by (Calzetti et al. 2000), assuming the stars and gas suffer the same reddening, and third, the same Calzetti et al. starburst extinction law, but instead, we assume the gas emission suffers higher extinction than the stars. The fitting algorithm is based on a traditional  $\chi^2$  approach described in Adamo et al. (2010a). The software also provide also error analyses as described in Adamo et al. (2012). Only sources that are detected with a photometric error  $\sigma \leq 0.3$  mag in at least four bands are analyzed. This condition will ensure that all the sources selected for visual inspection and a significant fraction of the excluded ones get a potential determination of their physical properties.

In the second case, the cluster physical properties are obtained using a Bayesian analysis method together with stochastically sampled cluster evolutionary models presented by Krumholz et al. (2015b, hereafter K15). The analysis is based on the Stochastically Lighting Up Galaxies (SLUG; da Silva et al. 2012; Krumholz et al. 2015a) code and its post-processing tool for analysis of star cluster properties, cluster\_SLUG (K15). The stellar libraries and extinction curves/attenuations used in cluster\_SLUG are the same as for Yggdrasil. In contrast to Yggdrasil, the SLUG treatment of the nebular emission is based on an analytical solution (see K15 for a discussion of the assumptions and possible differences arising from this different approach). cluster\_SLUG provides posterior probability distribution functions (PDFs) for the ages, masses, and extinctions of the cluster candidates, assuming different priors on the cluster mass function and dissolution rate.

This second approach provides a direct treatment of the uncertainties produced by low mass clusters affected by

stochasticity combined with stellar evolutions. In K15, we present a direct comparison of the recovered cluster properties using the two methods presented here.

Both deterministic and stochastic analyses are released, together with the photometric catalogs, on the LEGUS Web page. The data release for each galaxy contains 48 catalogs, 12 produced with deterministic fitting procedures and 36 with Bayesian approaches.<sup>45</sup>

Detailed studies on the impact of different flavors of stellar libraries and assumptions to build cluster evolutionary tracks are currently under investigation in the LEGUS team. Wofford et al. (2016) has tested the impacts of new evolutionary tracks in deriving the physical properties of very massive YSCs detected within a small sample of LEGUS galaxies. More specifically, we have tested the use of recently published stellar libraries with single non-rotating stars, i.e., Padova (Tang et al. 2014), Geneva (Ekström et al. 2012), Geneva with single rotating stars (Ekström et al. 2012), and BPASS with interacting binaries (Eldridge et al. 2008).

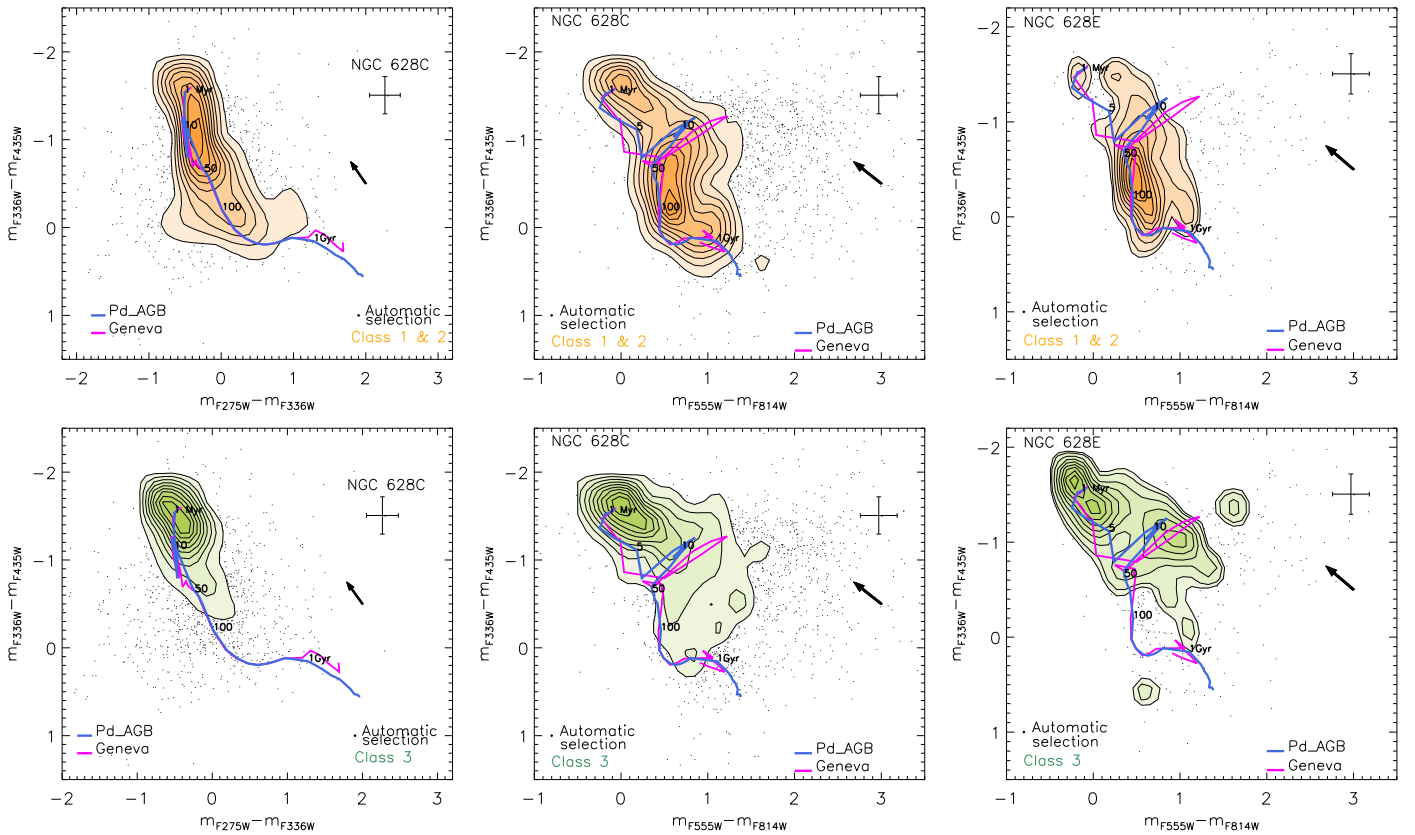
In the following sections, we will analyze the cluster population of NGC 628. The goal is to probe cluster formation conditions and evolution in this galaxy. We will check whether our results depend on the assumptions made to retrieve the cluster photometry and physical properties. Differences will be outlined and taken into account in our interpretations of the results.

#### 4. The Photometric Properties of the NGC 628 YSC Population

In the upcoming analysis, we use as a reference sample the YSC catalog of NGC 628 obtained with photometry corrected by an average aperture correction; SED fits produced with the deterministic approach, assuming Padova stellar evolutionary models and solar metallicity tracks; and the Milky Way extinction law (Cardelli et al. 1989) to take into account the internal reddening. The reference catalogs of NGC 628c and NGC 628e contain 3086 and 593 cluster candidates, respectively. Of these sources, roughly 1600 and 380 passed the selection required to move onto the next step of visual inspection (e.g., they have photometric errors  $< 0.3$  mag in all four bands and CI values  $> 1.4$ ). In NGC 628c (NGC 628e), 334 (92) systems have been classified as class 1, 357 (80) as class 2, and 326 (87) as class 3. The remaining 583 (121) sources have been assigned class 4.

The main difference between class 1 and class 2 clusters is linked to the apparent morphology of the systems, with the latter having elongated surface brightness. Ellipticity in the surface brightness profile of well-resolved nearby YSCs has been measured in the MW and the Local Group (e.g., San Roman et al. 2012), but they find that it is not linked to any specific dynamical status of the cluster (e.g., expanding and/or disrupting systems). In the LEGUS galaxies, clusters are not fully resolved and our ability to recover their morphology is very much dependent on the distance of the galaxy and the crowding of the region. In Grasha et al. (2015), we compare the cluster physical properties such as ages, masses, and

<sup>45</sup> Twelve catalogs come from a combination of the two photometric approaches for aperture correction (average based, CI based), two stellar libraries (Geneva and AGB-Padova), and three extinction/attenuation curves. The Bayesian analysis is produced for three different sets of priors but the same combination of photometric analysis, stellar libraries, and extinction curves ( $12 \times 3$  for a total of 36 catalogs).



**Figure 8.** UV (left) and optical (middle and right panels) color-color diagrams of the cluster population of NGC 628c (left and central) and NGC 628e (right). The dots show the location of all sources contained in the final reference (AV\_APCOR) catalog. The overlaid density number contours show the distribution of class 1 and 2 clusters (top), and class 3 objects (bottom). The contours include regions with densities equal to or larger than 10% of the 2D histogram peak. Padova (blue) and Geneva (magenta) evolutionary tracks are included (see the legend). The arrow shows in which direction the objects will move if corrected for a reddening  $E(B - V) = 0.2$  mag. The error bar on the top-right corner shows the average error in color.

extinctions of the three different classes (1, 2, 3) of systems identified in NGC 628. We report that the median age and mass of class 1 clusters are slightly higher than for the other two classes. This trend may be the result of the methodology we apply to classify cluster candidates. YSCs are born in cluster complexes and giant young star-forming regions. This implies that during the first  $\sim 10$  Myr, it is challenging to identify the real morphology of a cluster because of the complexity of the regions where clusters form (e.g., the Tarantula nebula in the LMC is a good nearby example). Conversely, we believe class 3 objects constitute a different type of system from class 1 and 2 objects. In particular, their multipeak nature suggests these to be likely compact associations. In light of these considerations, in the rest of the analysis we will look at cluster physical properties unifying class 1 and 2 objects under the same group and comparing their properties to class 3 objects.

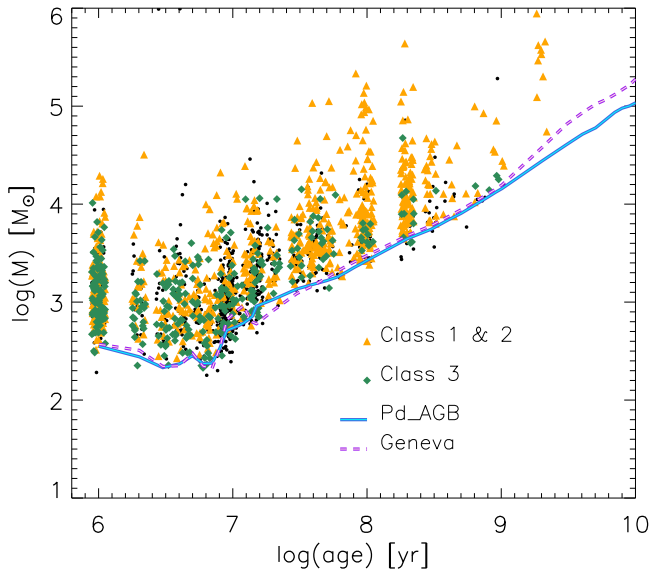
In Figure 8, we show the UV and optical color-color diagrams of all the sources included in the automatic catalog as dots, while the class 1 and 2 clusters and class 3 systems from our reference (AV\_APCOR) catalog are shown with contours of density numbers. The spread in colors of the automatically selected populations is clearly reduced after visual inspection. The color-color diagrams show that in the inner region of NGC 628, the cluster population is well distributed along the evolutionary track with the peak of number densities of clusters at very young ages ( $\sim 1$ – $10$  Myr) and between 50 and several hundreds of Myr. We see a clear difference in the distribution of class 1 and 2 clusters versus class 3; the latter is

fairly concentrated in a region of the diagram corresponding to 1–10 Myr and it extends up to  $\sim 50$  Myr, with relatively few objects at older ages. The difference between the age distributions of the classes of objects confirms that the morphological classification of class 3 objects, characterized by multiple peaks in their brightness profile and asymmetries in the light distribution is probably selecting stellar associations that evaporate and disappear in short timescales in the galactic stellar field (e.g., Gieles & Portegies Zwart 2011) and not gravitationally bound objects.

The age-mass diagram in Figure 9 confirms the observed trend in the color properties of our clusters and associations. The majority of class 3 objects have ages below 20 Myr and smaller masses than class 1 and 2 as already reported by Grasha et al. (2015). The maximum mass observed per age bin increases on average as a function of age, a result of the nearly constant SFR, the stochastic nature of the cluster formation process, and of the size-of-sample effect (longer age intervals imply a higher chance to form more massive clusters; e.g., Hunter et al. 2003).

We also investigate the colors of the cluster population detected in the outer pointing of NGC 628. The color-color diagrams reported in the panels on the right of Figure 8 show that the cluster population (class 1 and 2) in the outer region of NGC 628 has a less significant population of very young clusters (the contour levels are very weak in the region corresponding to ages younger than 5 Myr) with respect to the inner pointing. We see that the outer field color distribution is





**Figure 9.** Age-mass diagram of the NGC 628c pointing. The mass limits as a function of age for both evolutionary tracks, shown by the blue and purple lines, are estimated assuming  $M_V = -6$  mag (23.98 mag), i.e., the magnitude cut used to select objects for visual inspection. The underlying black dots show the sources in the reference catalog which have been detected in at least four filters but have not been visually inspected. A small random shift is applied to the age of each source so that they do not overlap, creating single columns at each age step. The discrete age steps of the models are anyway visible.

dominated by a cluster population with ages between 50 and a few hundreds Myr. On the other hand, the distribution of the class 3 associations behaves in a similar fashion in the two galactic regions, with their numbers quickly fading after 50 Myr. Similar behaviors in the YSC populations of the inner and outer regions have already been reported in the literature for M83 (e.g., Bastian et al. 2011; Chandar et al. 2014) and can likely be linked to the role of the environment where YSCs are forming and evolving. The large sampling of different galactic environments provided by LEGUS will be a key to interpret these observational trends in the near future.

#### 4.1. A Comparison Between the AV\_APCOR and CI\_BASED Catalogs of NGC 628c

In this section we compare the photometry and analysis performed on the CI\_BASED catalog to our reference catalog. For convenience, we show only the analysis performed on the NGC 628c pointing, but the outcome is very similar for the outer field.

The advantage of using average aperture corrections is that this method does not change the shape of the SEDs or the intrinsic colors of the clusters. As we see in Table 1, the differences between the applied aperture corrections change only slightly between different bandpasses (i.e., it reflects the change in the PSF as a function of filter/camera). Therefore, such a correction will change the normalization of the cluster SED, but not the shape. This method, however, does not take into account that, at the distance ranges of the LEGUS targets, clusters are partially resolved, that is, their CSF changes as a function of the cluster size. However, since the intrinsic shape of the SED is preserved and only the normalization is affected, this method may overestimate the mass of very compact clusters or underestimate the mass of the very extended

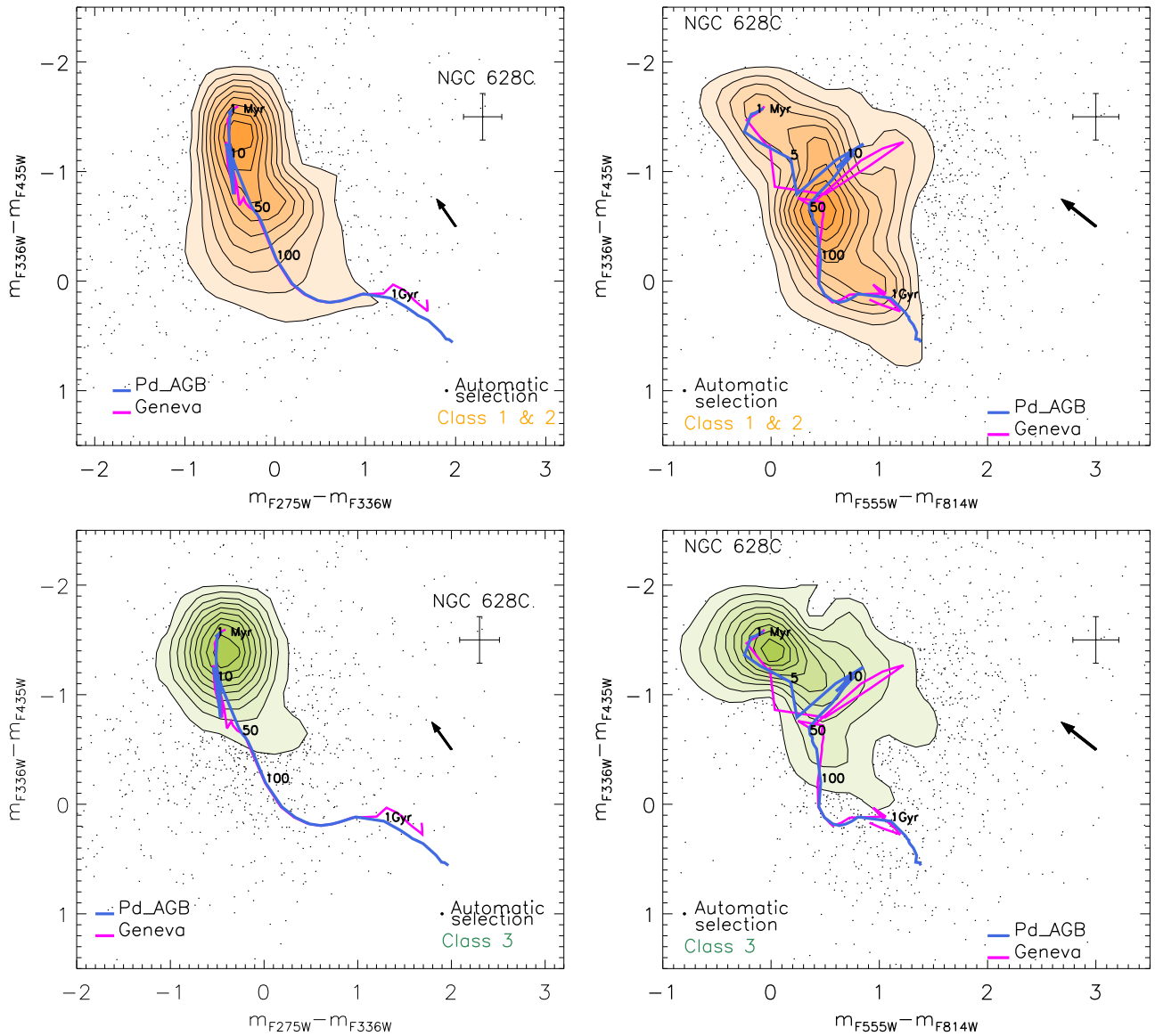
sources, but the uncertainties will be within the 0.1–0.2 dex in logarithmic age and mass usually produced by the fitting method.

The CI-based aperture correction takes into account the relation between the size of the cluster and the required aperture correction. The relation has been derived in each band using simulations of YSCs of varying sizes (see D. O. Cook et al. 2017, in preparation). This method has the great advantage of taking into account the cluster sizes. The limitation, however, resides in the uncertainties produced by the CI estimates as a function of wavelength. This means that the CI-based method not only changes the normalization of the SED but also the shape. Therefore, the uncertainties propagate in the estimated mass, age, and extinction of the source.

In Figure 10, we show the same UV and optical color-color diagrams but derived using the photometry of the CI\_BASED catalog. We notice that while the overall location of the clusters and associations are similar in the two photometric catalogs, the contours of the CI\_BASED photometry are more extended on the left side of the tracks, a spread that is larger than the photometric uncertainties. In general, one expects sources identified as clusters to diffuse on the right side of the evolutionary tracks because of reddening, while the spread on the left side is mainly produced by photometric errors and uncertainties in the calibration.

In Figure 11, the analysis of the residuals (i.e., the difference between the observed and the best model-integrated fluxes) of class 1, 2, and 3 clusters in each band does not show significant differences in the residuals of the F336W filter in both the AV\_APCOR and CI\_BASED catalogs. The residuals of the F275W filters in the CI\_BASED catalog are more concentrated around zero (the best match) than the AV\_APCOR ones. The opposite trend is observed in the AV\_APCOR residuals of the *BVI* filters, where they show a narrower distribution around the best match than the CI\_BASED ones. The latter are less peaked and show a tail toward positive values (i.e., the observed magnitude is brighter than the one predicted by the best model) in the *B* and *V* bands and negative values in the *I* band. Overall, we see that the differences between the reduced  $\chi^2$  obtained from the fit performed to the photometry of the AV\_APCOR and CI\_BASED catalogs show a more significant negative tail, suggesting that the fit to the CI\_BASED photometry is slightly worse.

A direct comparison of the recovered ages, masses, and extinctions of class 1, 2, and 3 objects of the AV\_APCOR and CI\_BASED catalogs is shown in Figure 12. Although on average we see correspondence along the one-to-one line, some deviations are also significant. The most important difference in the CI\_BASED catalog is the appearance of the pronounced peak at around 5 Myr, visible in the top-right distribution of Figure 12. According to the top-left panel, these systems have been assigned in the AV\_APCOR catalog both younger, similar, or older ages. Since the mass-to-light ratio is smaller at younger ages, the objects that become younger in the CI\_BASED analysis will also have smaller stellar mass; therefore, we can explain why the AV\_APCOR and CI\_BASED mass histograms (central right panel of Figure 12) differ in the mass bins between 1000 and  $10^4 M_\odot$ . In the CI\_BASED catalog, these objects have been assigned masses  $\leq 10^3 M_\odot$ . In total, we estimate that about 20% of the class 1, 2, and 3 systems that are in common in the two catalogs have differences in ages larger than 0.1 dex, which is the average



**Figure 10.** UV (left) and optical (right) color-color diagrams of the cluster population of NGC 628c based on CI\_BASED catalogs. The dots show the location of all the sources contained in the CI\_BASED catalog. The overlaid density number contours show the distribution of class 1 and 2 clusters (top), and class 3 objects (bottom). See Figure 8 for more details.

uncertainty recovered for age estimates obtained with deterministic methods. This small fraction of deviating sources may explain why we do not see significant differences in the recovered CLFs, CMFs, and disruption rates (Section 5) of the cluster population in NGC 628 when AV\_APCOR and CI\_BASED catalogs are used. Further investigation of these two approaches is presented in D. O. Cook et al. (2017, in preparation). In the next section, we will report the results of the analysis performed on the YSC population of NGC 628 using only the AV\_APCOR reference catalog.

## 5. Constraining the Formation and Evolution of Clusters and Associations in NGC 628

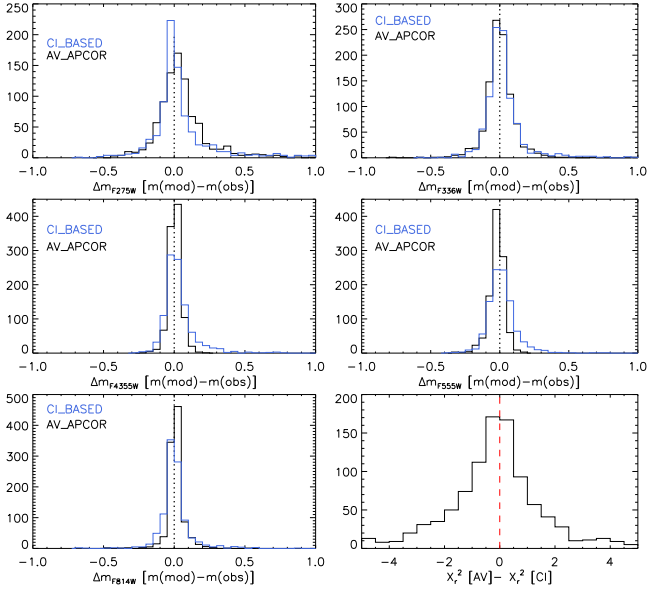
In Section 4, we have observed that the photometric properties of clusters (class 1 and 2, in our analysis) and compact associations (class 3) show significant differences, with the latter class disappearing from the regions of the color-color diagrams occupied by more evolved stellar populations.

In the following sections, we analyze and compare physical and statistical properties of class 1 and 2 and class 3 objects separately. The aim is to probe differences and analogies between these two types of stellar objects that can help us to understand their formation process and evolution.

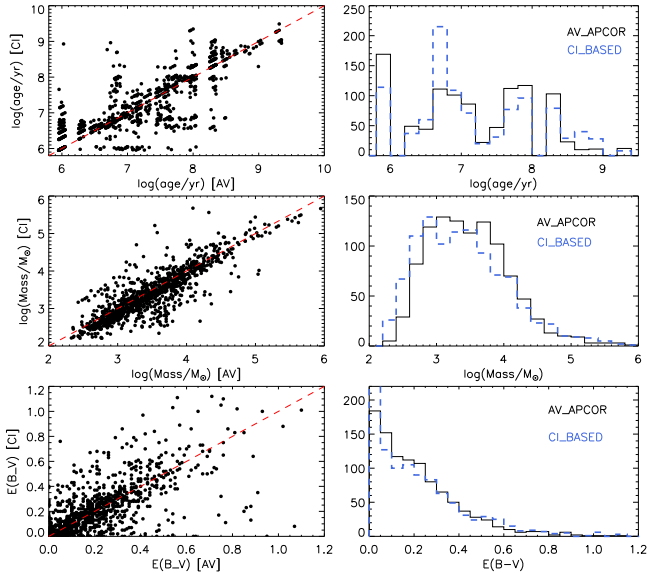
### 5.1. Multiwavelength Analysis of the CLF of Likely Bound and Unbound Systems

The luminosity function of YSCs is typically described as a power-law function in the luminosity space,  $dN/dL \propto L^{-\beta}$ . However, we fit the function in logarithmic space, where the luminosity is replaced by the magnitude, so that  $\log[d(N)/dM] \propto \theta \times M$ , where  $\beta = 2.5 \times \theta + 1$  (e.g., Whitmore et al. 2002; Haas et al. 2008).

We apply two different techniques, equally used in the literature, to analyze the CLF of NGC 628. In Figure 13 we report the observed luminosity functions in each band, using cumulative distributions (left panel), like in Bastian et al. (2012),



**Figure 11.** Residuals produced by the SED fitting analysis of both AV\_APCOR (black) and CL\_BASED (blue) catalogs of NGC 628c as a function of wavelength. The bottom-right panel shows the difference between the recovered reduced  $\chi^2$  in the AV\_APCOR and CL\_BASED catalogs. Only class 1, 2, and 3 objects have been used to produce these distributions.



**Figure 12.** Recovered physical properties for class 1, 2, and 3 objects for both AV\_APCOR (x-axes and black solid line histogram) and CL\_BASED (y-axes and blue dashed line histograms) catalogs of NGC 628c. The dashed red lines in the left panels show the location of the one-to-one correlation.

and bins containing the same number of objects (right panel), following the Maíz Apellániz & Úbeda (2005) approach.

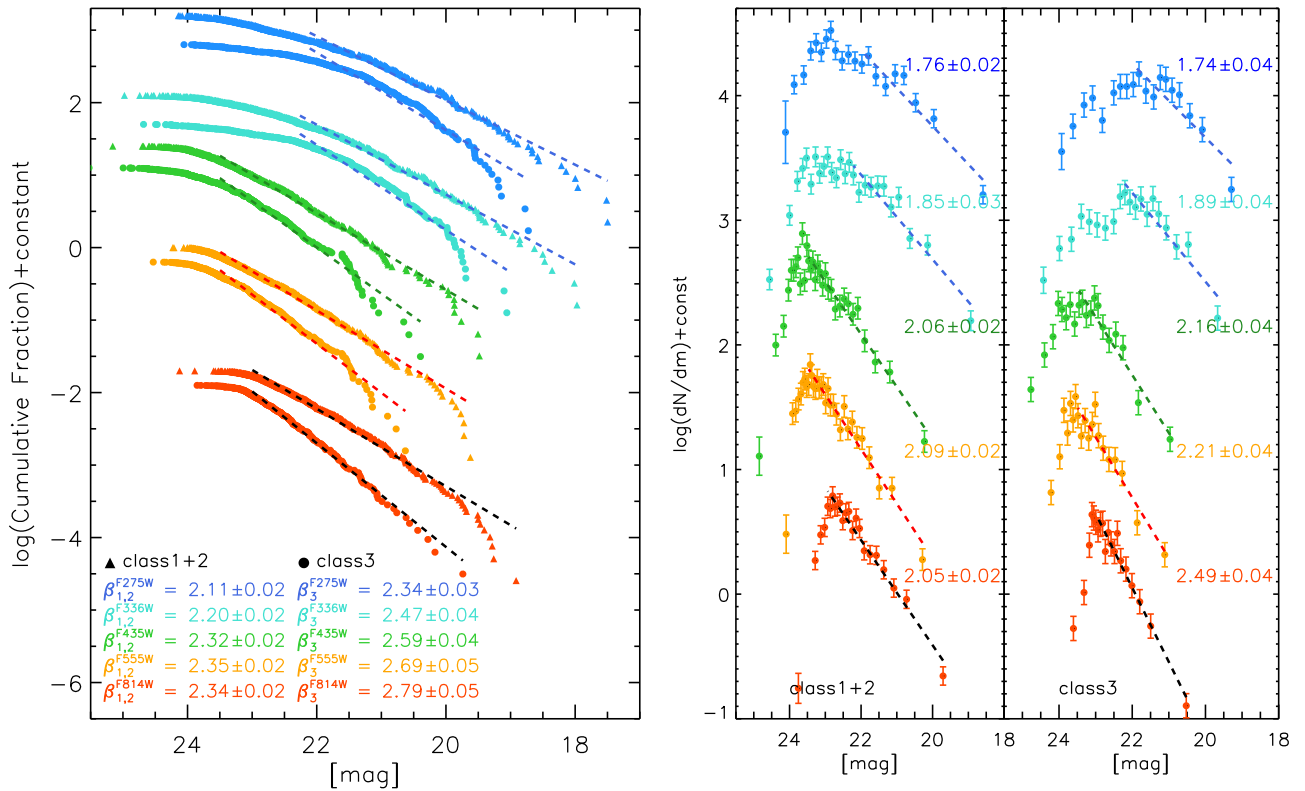
The luminosity properties of the clusters are directly observable, not affected by age or mass determinations. However, their luminosity distributions depend on both the detection limits of our data sets and by the adopted extraction procedure combined with the selection criteria we impose to yield the final catalog. To build the CLF, we select only sources that have been visually classified as class 1, 2, or 3 (thus they are brighter than  $-6.0$  mag in the  $V$  band and detected in four filters with photometric error smaller than  $0.3$  mag) and are

younger than 200 Myr. In Figure 14, we show the age–mass diagnostic diagram including the recovered 90% detection limits in the four bands required for detection and the  $V$ -band cut at  $-6.0$  mag, all converted to limiting masses as a function of age. We observe that the  $V$ -band cut applied for the visual inspection is more conservative than the detection limits of our data set. However, we notice here that the resulting flattening of the distributions at the low luminosity ends is produced by the combination of a sharp magnitude cut combined with detection limits of the science frames and the method used to produce the final position catalog of cluster candidates. The age limit of 200 Myr enables us to directly compare the CLF to the CMF (see Section 5.3). In total, we count 733 (370) class 1 and 2 (class 3 numbers are indicated inside brackets) objects in the F275W filter, 846 (404) in the  $U$  band, and 851 (408) in the  $BVI$  bands before the age cut. After the age cut is applied, we are left with 703 (369), 778 (397), and 783 (401) class 1 and 2 (class 3) objects in the  $UV$ ,  $U$ , and  $BVI$  bands, respectively. To prevent incompleteness from affecting our analysis, we perform the fit of the binned and cumulative distributions from the brightest bin (object) down to the bin (object) with magnitude brighter than 22.12, 22.26, 23.50, 23.50, and 23.0 mag in  $UV$ ,  $U$ ,  $B$ ,  $V$ , and  $I$ , respectively. These limiting values have been chosen to avoid the shallower regions of the distributions and are more conservative than the 90% completeness magnitudes reported in Table 1.

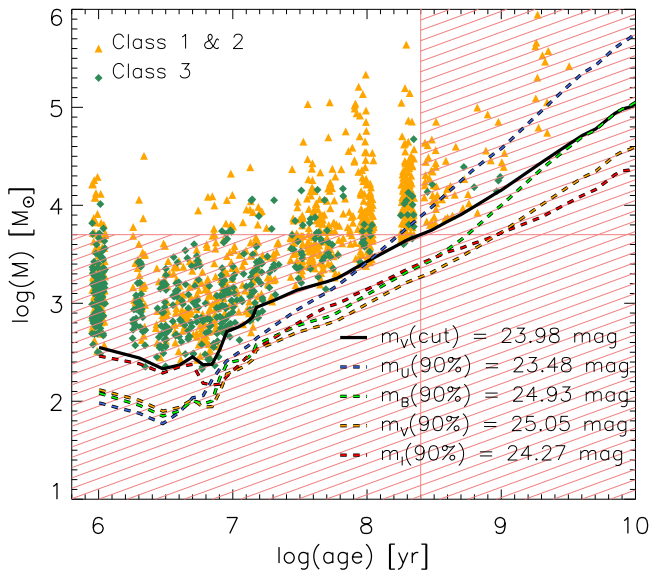
In the case of an equal number of object bins, the slopes and the associated uncertainties are produced by the IDL package LINFIT, which takes into account the weighted error associated with each bin. In the case of the cumulative functions, the error analysis has been performed with bootstrapping techniques. To take into account how the photometric uncertainty associated with each point affects the final recovered slope, we perform Monte Carlo realizations of 1000 cumulative distributions. With each observed magnitude, we associate an uncertainty extracted from a Gaussian distribution with standard deviation equal to the maximum tolerated error of  $0.3$  mag. Each cumulative realization is thus fitted. The final error associated with the observed slope is the standard deviation of all the recovered indexes.

In general, both methods produce slopes that are close to an index of  $-2$ . We notice that the recovered slopes for the binned data are shallower than in the case of cumulative distributions. This is mainly the result of the differences between the two techniques (see Section 5.3), hence slopes determined using binned and cumulative distributions cannot be directly compared. Some important features are, however, observable in both analyses. We find a clear steepening as a function of increasing wavelength, i.e., the recovered slopes become significantly steeper than  $-2$  in the  $BVI$  filters. This is true for both class 1 and 2 clusters and class 3 associations (see insets in Figure 13). The distributions of the two bluest filters show an extended flat peak that in the cumulative distributions appears as a significant curvature. Moreover, the cumulative distributions at all wavelengths show a clear steepening at the bright magnitude ends. Such steepening is not observed in the binned distributions because while a variable-size binning technique mitigates biases introduced by equally spaced binning approaches (Maíz Apellániz & Úbeda 2005), it also tends to wash out small-scale variations (see Section 5.3). Because of the small number of very luminous objects, the brightest bin of the CLF has a width of about  $1.5$ – $2$  mag





**Figure 13.** Luminosity function of the whole cluster population of NGC 628 in the five standard LEGUS bands. From top to bottom, we plot the CLFs obtained in the *UVUBVI* filters. On the left panel, we fit cumulative distributions of the magnitudes of class 1 and 2 and class 3, the recovered slopes are listed in the insets. On the right plot we fit distributions of bins containing the same number of objects. The plot consists of two panels, the left one shows the distributions for class 1 and 2, the right one for class 3. The fit has been performed including the brightest object (bin) down to the system (bin) with magnitude comparable to the detection limits listed in Table 1.



**Figure 14.** Age-mass diagram of the cluster and association populations of both the inner and outer pointing of NGC 628. The magnitudes corresponding to 90% completeness limits (see Table 1) in the four bands required for the analysis have been converted into mass limits as a function of age using Yggdrasil models. We also include the detection limits in age and mass imposed by the  $M_V < -6 \text{ mag}$  selection criterion. The latter cut ensures we are above 90% recovery in all four bands for masses above  $5000 M_{\odot}$  and ages up to about 200 Myr. The shadowed areas show which part of the sample has been excluded in the analysis of the CMF and disruption rates.

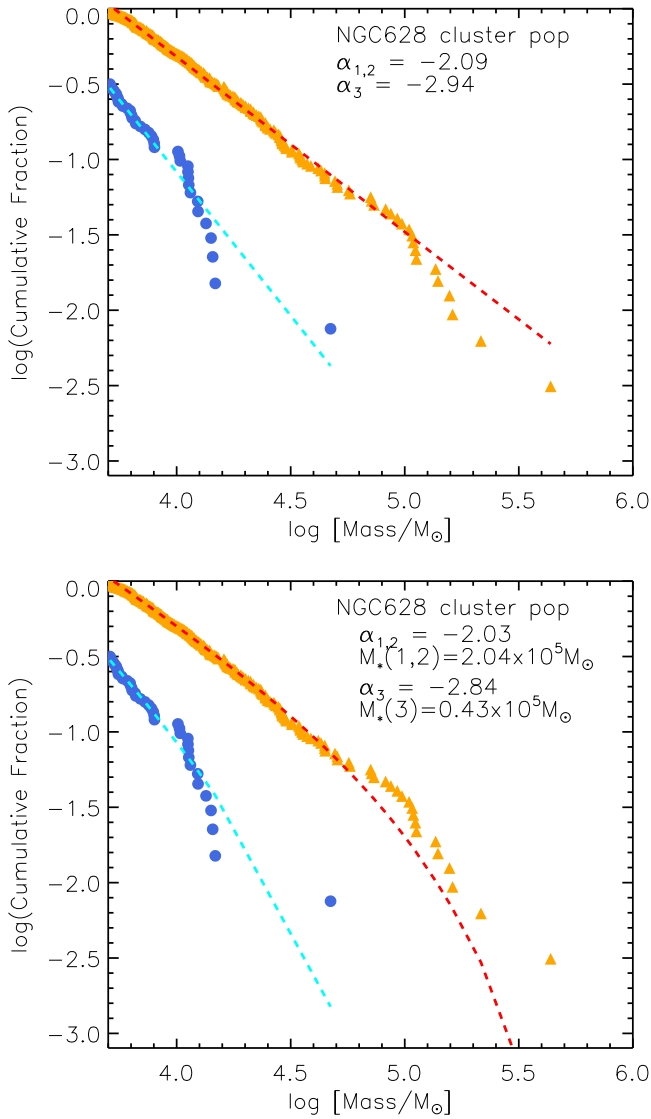
encompassing the range where the steepening in the cumulative function is observed. Our simulations, in Section 5.3, show that the departures from a single power-law function happen mainly

at the bright (massive) end, thus the cumulative analysis is better suited to investigate the shape of the luminosity and mass functions.

The luminosity function is a direct observable of the underlying mass function integrated over time. The trends we observe in the CLF of NGC 628 suggest a dearth of luminous (massive) clusters/associations and thus simultaneously analyzing the CLF and CMF can tell us something important about how clusters form and evolve in this galaxy.

## 5.2. The Analysis of the Cluster Mass Function

To derive masses, we need to perform an extra step where the ages and internal reddening of the sources are extracted. As discussed in K15, the stellar physical properties derived with our deterministic method are severely biased at very low masses due to stochastic variations from the small number of stars. Figure 14 of K15 shows the one-to-one comparison between Yggdrasil deterministic and SLUG stochastically derived cluster properties, suggesting that important deviations occur at cluster masses below  $5000 M_{\odot}$ . This mass limit was, in recent years, widely adopted in the YSC analysis based on deterministic approaches. We assume the same mass limit in our current analysis. Figure 14 shows the age-mass diagram of the clusters/associations in our two *HST* pointings of NGC 628. A mass cut of  $5000 M_{\odot}$  gives us complete detection up to a stellar age of 200 Myr. Sources falling within the shadowed areas are, thus, not included in the analysis of the CMF and disruption rates presented hereafter. In total, we count 320 class 1 and 2 clusters and 42 class 3 objects that pass this mass and age selection. We notice that our mass cut is very close to the



**Figure 15.** Cumulative mass functions of class 1 and 2 (orange triangles) and class 3 (blue dots) systems. The distributions have been created only with objects younger than 200 Myr, and the fit includes only systems more massive than  $5000 M_{\odot}$ . The recovered slopes for the two subpopulations are reported in the inset.

completeness limits of our catalogs at the last age bin at 200 Myr. We tested whether using a higher mass cut at  $8000 M_{\odot}$  produces different outcomes. We do not see any change in the recovered CMF properties, only higher errors because of the smaller number of clusters available for the analysis.

The observed mass function of class 1 and 2 (orange triangles) and class 3 (blue dots) systems are shown as cumulative distributions in Figure 15. The observed cumulative distributions are fitted using the IDL maximum-likelihood fitting package `MSPECFIT` (Rosolowsky 2005). We perform two different fits to the cumulative distributions, a single power-law function, in the integral form  $N(M' > M) \propto M^{\alpha}$ , and a power-law function with a truncation at the upper mass end, i.e.,  $N(M' > M) \propto N_0[(M/M_{*})^{\alpha} - 1]$  (see Rosolowsky 2005 and references therein for a complete discussion of the formalism). The resulting fitted parameters for the two functions (single power law in the top panel, truncated power law in the bottom panel) are included in the insets of Figure 15. In Table 2 we list the recovered values for the class 1 and 2

population.  $M_{*}$ , the index  $\alpha_{\text{SF}}$ , and  $N_0$ , which is the number of objects more massive than  $2^{1/\alpha} M_{*}$ , are determined for a truncated function, while the pure power-law fit provides the index  $\alpha_{\text{PLF}}$ . As described in Rosolowsky (2005), if the resulting  $N_0$  is significantly larger than 1, then a truncated CMF form is preferred to the more traditional single power-law function. When  $N_0 < 1$ , the truncation mass is unconstrained and thus a single power-law fit is sufficient. In Table 2, we also include errors. The errors associated with the observed maximum cluster mass,  $M_{\text{max}}$ , and fifth most massive cluster mass,  $M_{\text{max}}^{\text{5th}}$ , have been computed during the SED fitting procedure and described in Section 3. The errors associated with the best-fitting parameters have been computed using deviations from 1000 iterations of bootstrap trials.

In general, we observe that both a very steep single power-law fit and a truncated function fit with a slightly flatter index can reproduce the observed mass distribution for class 3 objects. However, the number of associations is very small (42), and thus it is not possible to impose any further constraint.

On the other hand, the analysis of the mass distributions of class 1 and 2 systems yields, for both a single power law and a truncated function type fits, slopes very close to  $-2$ . However, as already noticed during the analysis of the CLF, the approximation of the CMF by a single power-law function (see the top panel of Figure 15) overestimates the expected number of clusters at the upper mass end of the distribution. A fit to the observed CMF of class 1 and 2 with a truncated power-law function (bottom panel of Figure 15) yields a similar slope, but it mitigates the differences at the high mass end of the CMF distribution. The resulting  $N_0$  (see value listed in Table 2) is larger than 1, suggesting that the latter function provides a better fit to the observed CMF. Thus, a truncated function with slope  $\alpha_{\text{SF}} = -2.03$  and  $M_{*} \sim 2.0 \times 10^5 M_{\odot}$  is the statistically favored description of the observed CMF of NGC 628. However, it is important to notice that the number of clusters more massive than  $5 \times 10^4 M_{\odot}$  is about 22 and only half of those clusters are more massive than  $10^5 M_{\odot}$  so the constraint on  $M_{*}$  is weak and the uncertainties on  $N_0$  large.

As an exercise, we try to estimate the expected number of clusters more massive than  $M_{*}$ . Using the combination of far-UV and  $24 \mu\text{m}$  fluxes of the area covered by the LEGUS pointings of NGC 628, we estimate an SFR of about  $0.59 M_{\odot} \text{ yr}^{-1}$ . Assuming that the SFR was constant for the last 200 Myr, we estimate that a total stellar mass of  $1.18 \times 10^8 M_{\odot}$  has been formed in the region. Using the cluster formation efficiency definition given in Adamo et al. (2015) and clusters in the age range between 1 and 100 Myr (same as the age range to which the estimated SFR is sensitive to), we derive for this region of NGC 628 a cluster formation efficiency of 12%. This means that 12% of the total stellar mass of  $1.18 \times 10^8 M_{\odot}$  is in bound clusters, i.e.,  $1.42 \times 10^7 M_{\odot}$ . Using the latter amount as the total stellar mass in clusters, we can estimate the number of clusters more massive than  $M_{*}$ . Observationally, we find two clusters more massive than  $M_{*}$ . Assuming a pure power-law mass function of slope  $-2.09$  (with upper mass  $1. \times 10^7 M_{\odot}$ ), we estimate that five clusters more massive than  $M_{*}$  should have formed in the last 200 Myr. A Schechter-type function, as described by Equation (3), results in one cluster more massive than  $M_{*}$ . The estimated total stellar mass in clusters results in cluster numbers that are consistent with the observed ones but does not produce any definitive proof that can help to discern the real shape of the upper mass function. Therefore, the

**Table 2**  
Parameters Describing the Cluster Mass Function in NGC 628<sup>a</sup>

Region	Number	Radius (kpc)	$M_{\max}$ ( $10^5 M_{\odot}$ )	$M_{\max}^{5th}$ ( $10^5 M_{\odot}$ )	$N_0^b$	$\alpha_{SF}^{b,c}$	$M_{\star}^b$ ( $10^5 M_{\odot}$ )	$\alpha_{PLF}^c$
bin 1	107	0.46–3.19	$1.62^{+0.13}_{-0.26}$	$1.11^{+0.07}_{-0.30}$	$2.29 \pm 3.40$	$-1.84 \pm 0.13$	$4.85 \pm 2.02$	$-1.90 \pm 0.10$
bin 2	107	3.22–4.53	$2.16^{+0.18}_{-0.19}$	$0.45^{+0.11}_{-0.10}$	$3.77 \pm 2.80$	$-2.15 \pm 0.12$	$0.98 \pm 0.42$	$-2.25 \pm 0.12$
bin 3	106	4.56–10.15	$4.36^{+0.14}_{-0.27}$	$0.43^{+0.04}_{-0.11}$	$5.67 \pm 4.17$	$-2.00 \pm 0.13$	$1.04 \pm 0.59$	$-2.13 \pm 0.09$
all	320	0.46–10.15	$4.36^{+0.14}_{-0.27}$	$1.40^{+0.09}_{-0.08}$	$7.58 \pm 4.20$	$-2.03 \pm 0.07$	$2.03 \pm 0.81$	$-2.09 \pm 0.06$

**Notes.**

<sup>a</sup> The table only includes observed and fitted values of class 1 and 2 cluster population with ages smaller than 200 Myr and masses above 5000  $M_{\odot}$ .

<sup>b</sup> Mass function parameter fits, computed via the maximum-likelihood method of Rosolowsky (2005). If  $N_0 \gg 1$ , a truncated CMF form is appropriate, while  $N_0 \lesssim 1$  indicates a single power law is more appropriate.

<sup>c</sup>  $\alpha_{SF}$  is the slope derived assuming a truncated mass function, and  $\alpha_{PLF}$  has been derived assuming a pure power-law function. See the text for details.

solution provided by a pure power-law function with slope  $-2.09$  cannot be discarded.

We also test whether there is any variation in the CMF properties as a function of galactocentric distance by performing a radial analysis of the CMF for the class 1 and 2 clusters. In total, our sample contains 320 objects more massive than 5000  $M_{\odot}$  and younger than 200 Myr. We divide the sample in three radial bins of increasing distance from the center of the galaxy, each containing the same number of objects so that we remove the size-of-sample effect. We then determine the  $M_{\star}$ ,  $N_0$ ,  $\alpha_{SF}$ , and  $\alpha_{PLF}$  of the mass function of each bin using both a truncated and a more traditional power-law function as described above. The recovered values are listed in Table 2. In bin 1, the recovered  $M_{\star}$  is larger than the most massive cluster observed in the bin, and the resulting  $N_0$ , including the uncertainties, is very close to 1. This means that the shape of the upper mass end of the CMF in the inner bin remains unconstrained and the solution with a single power-law fit is as likely. In the second and third bins, between 3 and 10 kpc,  $M_{\star}$  is consistently  $1 \times 10^5 M_{\odot}$  and does not decline significantly. The recovered  $N_0$  is larger than 1 but the uncertainties are large, too. In Figure 16, we plot the derived  $M_{\star}$  (cyan dots and blue bands, including the uncertainties), the masses and uncertainties of the most massive cluster,  $M_{\max}$  (red triangles and orange bands), and the fifth most massive cluster,  $M_{\max}^{5th}$  (green triangles and bands), observed in each bin. The derived  $M_{\star}$  decreases significantly between the innermost and the other two bins. In the two outermost bins, the observed masses of the most massive clusters are significantly more massive than the constrained  $M_{\star}$ , but the numbers of clusters close to the determined  $M_{\star}$  are small, which makes the statistics (e.g.,  $N_0$ ) quite uncertain.

Overall, we observe that if the recovered  $N_0$  is larger than 1, the analyzed cluster population has formed at least a few clusters with masses close to and larger than the truncation value,  $M_{\star}$ . This behavior suggests that  $M_{\star}$  is not a sharp truncation and that the mass function is likely stochastically sampled. In the literature, it has been suggested that the CMF of YSC populations in local galaxies can be described with a Schechter function of slope close to  $-2$  and a rapid exponential decline above a certain truncation mass (Gieles et al. 2006; Larsen 2009; Bastian et al. 2012). The Schechter function differs from a pure power-law function only at the upper mass end of the distributions and could in principle explain the disagreement between the low numbers of observed massive clusters with respect to the expected one from the extrapolation

of a pure power-law function. The probability to form clusters more massive than  $M_{\star}$  declines exponentially but is not null.

We notice that the  $M_{\star}$  of class 1 and 2 systems recovered for NGC 628 is very similar to the one retrieved for M83 (Adamo et al. 2015), in agreement with the evidence presented in Larsen (2009), who suggested that  $M_{\star}$  in local spiral galaxies is about a few times  $10^5 M_{\odot}$ .

### 5.3. Using Monte Carlo Simulations to Link Cluster Mass and Luminosity Functions

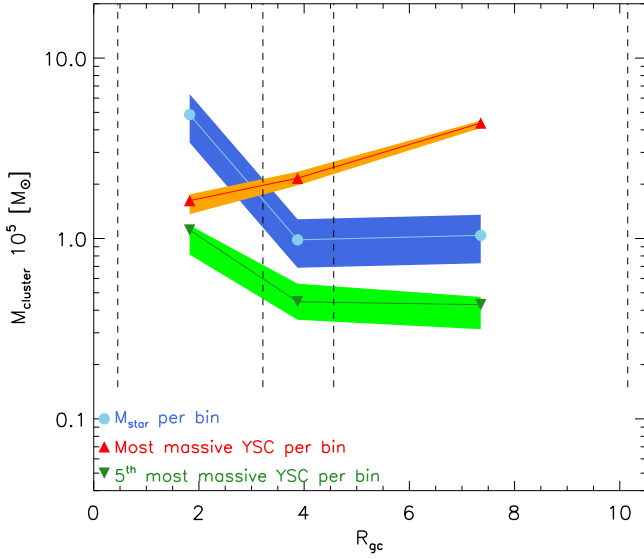
In the attempt to understand the uncertainties imposed by the low number statistics and, at the same time, link the observed CMF to the CLF we use simulated cluster populations. We stochastically sample a pure power-law mass function (we will refer to this population as run A) and a truncated one in the form of a Schechter function (run B), with the same slopes and  $M_{\star}$  derived for the cluster (class 1 and 2) population of NGC 628. In both runs, we sample the mass function from  $2 \times 10^2$  to  $100 \times M_{\max} M_{\odot}$ . We use  $1.5 \times 10^4$  objects so that the resulting cluster population will approximately have the same number of clusters more massive than 5000  $M_{\odot}$  as observed in NGC 628 (i.e., 320 class 1 and 2 objects). For each cluster, we stochastically assign an age between 1 and 200 Myr, i.e., we assume that star formation has been constant during this time range. Based on the evidence produced by the analysis of cluster dissolution timescales in Section 5.4, we also include mass-dependent disruption using the formula:

$$M_i(t_i) = (M_i^{\gamma} - \gamma t_i/t_0)^{-\gamma}, \quad (1)$$

where we assume  $\gamma = 0.65$  and  $t_0 = 4.91 \times 10^5$  years (values obtained from a maximum-likelihood fit to the data; see Section 5.4 for details). We then create 1000 realizations of each population.

In Figure 17, we include the observed CMF of NGC 628 (orange dots) and we overplot the median (red solid line) mass cumulative distribution and the distributions containing 50% (dashed red line) and 90% (dotted red line) of the 1000 Monte Carlo realizations sampled from an underlying pure power-law mass function (run A, top panel) and Schechter truncated function (run B, bottom panel). To build the cumulative distributions of the simulated populations we apply the same mass cut as in the observations at 5000  $M_{\odot}$ . When comparing the loci occupied by the simulations with respect to the observed CMF we see that 90% of realizations of run A overestimate the number of clusters more massive than  $10^5 M_{\odot}$ .





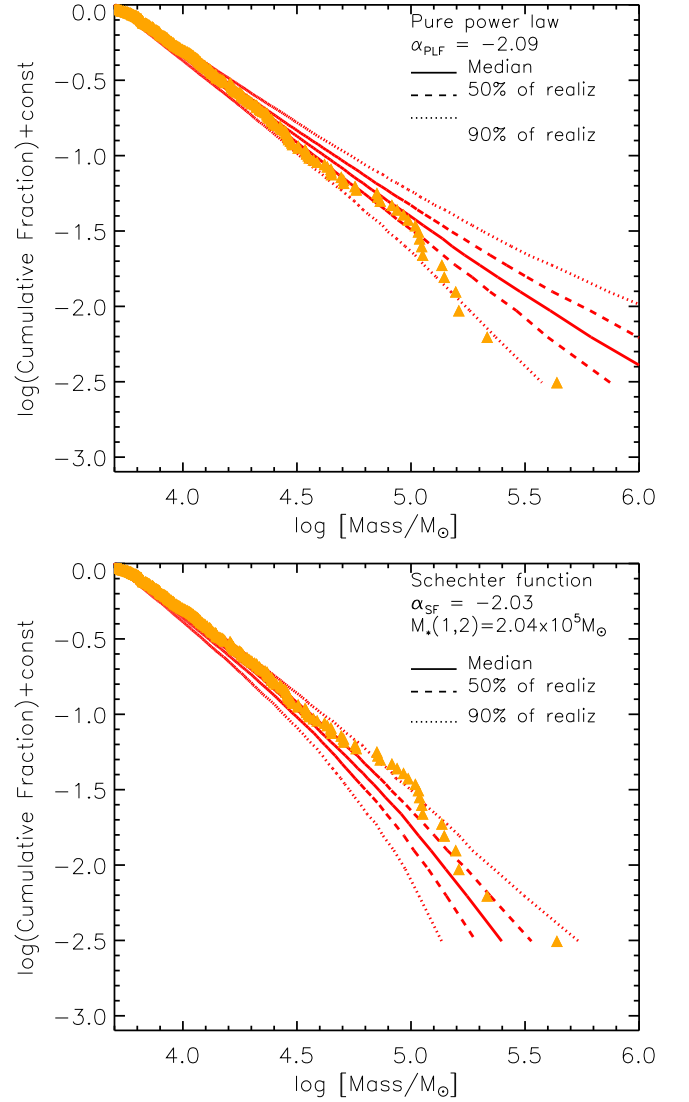
**Figure 16.** Comparison between the estimated truncation mass,  $M_*$  (cyan dots), and the observed mass of the most massive cluster (red triangles) and fifth most massive cluster (green downward triangles) in bins of increasing galactocentric distance. The vertical dashed lines show the width of the bins, which were created to contain the same total number of class 1+2. The shadowed areas show the uncertainties associated with each derived  $M_*$  and the errors on the mass estimates obtained with the SED fitting procedure described in Section 3.

(the locations of the upper mass end of the observed CMF is off the 90% limits of realizations). We test the null hypothesis that the upper part of the median distribution of run A and run B realizations and of the observed CMF are drawn from the same parent population. Since the differences between the pure power-law function and the truncated function is at the upper mass end of the distributions, we run both a Kolmogorov–Smirnov (KS) and an Anderson–Darling (AD) test using clusters more massive than  $10^4 M_\odot$ . We recover similar probabilities from both statistics with  $p(\text{AD})$  and  $p(\text{KS})$  of  $\sim 0.3$  and  $\sim 0.9$  for run A and run B, respectively. This result does not discard any of the two functions, but it yields a marginal preference for a truncated mass function.

Next, we analyze the resulting luminosity functions produced from the two run A and run B simulated cluster populations with the goal of understanding if the underlying CMF has a truncation. We use the median cluster population of the 1000 Monte Carlo realizations to yield the CLF. We do not apply any mass cut to the simulated populations, but we use the age and mass of each mock object together with Yggdrasil models to estimate their fluxes in the *UBVILEGUS* bands. We then select only sources with a luminosity brighter than  $M_V \leq -6$  mag and brighter than the 90% completeness values in the other two bands for which we require detection for a source to enter the catalog (namely *B* and *I*). Extinction is not taken into account.

The resulting cumulative and binned CLFs as a function of waveband of the two simulated cluster populations are illustrated in Figure 18. As done in Section 5.1 for the observed CLFs, we fit a single power-law function to the simulated CLFs. To recreate the limiting detection depth of each waveband, we use the same low luminosity limits as for the observed distributions.

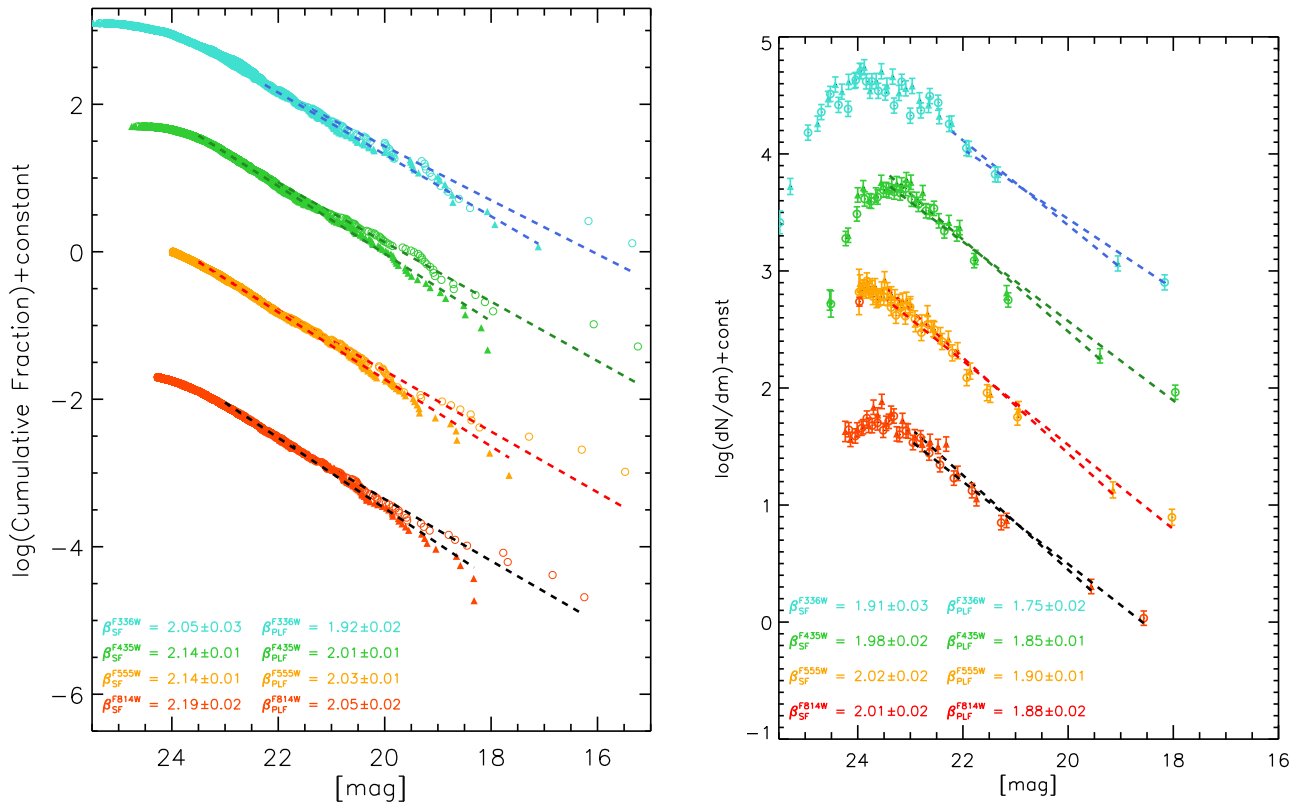
Since the underlying CMF of the simulated cluster populations is known, we can directly verify the effect of using either binning or cumulative techniques on the resulting



**Figure 17.** Monte Carlo simulations aim to reproduce the observed CMF for class 1 and 2 (orange triangles). 1000 realizations, containing a similar number of objects as in the real distributions, have been performed in each case. The median and the regions containing 50% and 90% realizations are shown as indicated in the inset. In the top panel, we assume a pure power-law shape (run A). In the bottom panel, we assume a Schechter-type distribution (run B). The slopes of the power law and the truncation masses used are reported in the top-right inset of each panel. See the text for a discussion of the results.

distributions. We notice that the recovered slopes  $\beta_{\text{PLF}}^\lambda$  and  $\beta_{\text{SF}}^\lambda$  listed in the left panel of Figure 18 are about 0.2 steeper than the ones listed in the right plot of the same figure. As discussed in Section 5.2, the way the equal-number object binning and cumulative distributions are built make them sensitive to different properties of the distributions. Since we are interested in understanding whether the CMF has a truncation at high mass, we prefer to analyze cumulative distributions.

The cumulative distributions in the left panel of Figure 18 should be compared to the observed cumulative distributions of class 1 and 2 objects (cumulative functions illustrated with triangles in the left plot in Figure 13). First, the resulting indexes of run A ( $\beta_{\text{PLF}}^\lambda$ ) and run B ( $\beta_{\text{SF}}^\lambda$ ) show that the effect of fading when comparing the slope of the *U* band to the ones in the redder filters is between 0.1 and 0.15. The differences in the



**Figure 18.** Cumulative function (left) and binned distributions (right) of the luminosity produced with run A (pure power-law function, open circles) and run B (Schechter truncated function, solid triangles) simulated cluster populations. Dashed lines are the single power-law fits performed on both distributions. The fit is carried out down to the same luminosity limits as in the observed CLFs to take into account the effect of detection limits imposed by the depth of the data. The resulting indexes for run A ( $\alpha_{\text{PLF}}$ ) and run B ( $\alpha_{\text{SF}}$ ) are listed in the bottom-left insets.

recovered slopes for the observed CLFs in the ultraviolet and optical are larger (about 0.2–0.3). However, we note that our simulations do not include extinction, which preferentially affects the bluer bands and could explain the larger flattening of the observed ultraviolet slopes.

Second, we notice that the recovered CLF slopes for a population of clusters sampled out of a pure power-law CMF (run A) are very close to the initial slope of the corresponding CMF (i.e.,  $\alpha_{\text{PLF}} = 2.09$ ). The slopes obtained fitting a CLF sampled out of a CMF with a power-law shape (slope  $-2$ ) and a truncation at the high mass end (run B) are steeper than  $-2$ . We also notice a fundamental difference at the bright luminosity end between the cumulative distributions of the two runs (pure power law versus power law with a truncation). The presence of a truncation at the high mass end of the CMF (run B) results in significant deviations at the bright end (i.e., solid triangles) of the CLF, with respect to the population sampled out from a pure power-law CMF (run A, open circles). The drop at the bright luminosity observed for the run B CLF (filled circles, left panel of Figure 18) is very similar to the ones observed in the real CLF (class 1 and 2, left plot in Figure 13). This trend reinforces the evidence of a presence of a truncation in the underlying CMF of NGC 628.

#### 5.4. The Cluster and Association Age Distributions and Implications for Evolution

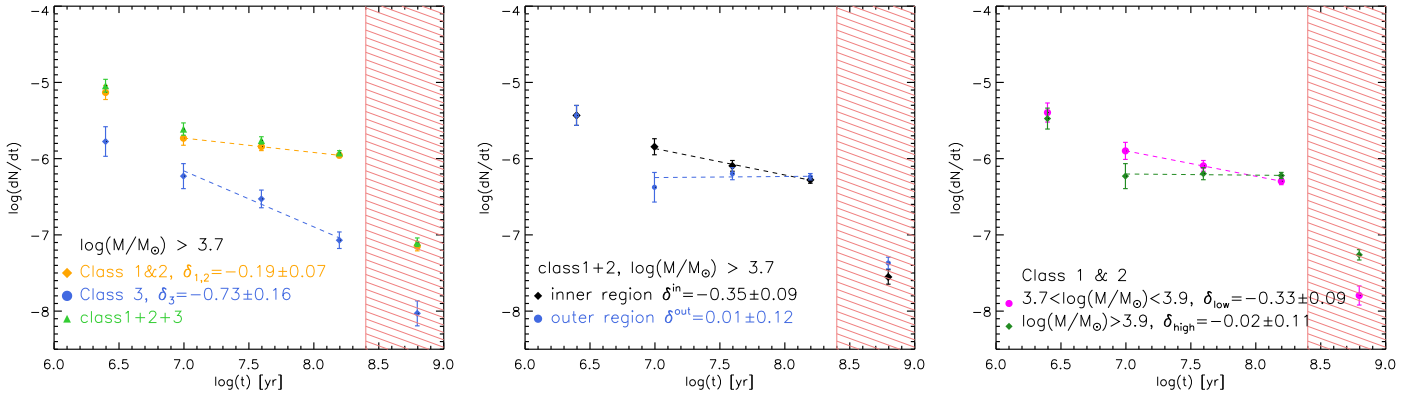
In this section, we probe the evolution of the two classes of objects visually identified: class 1 and 2 likely being clusters,

and class 3 being a heterogeneous sample dominated by compact stellar associations.

We apply the same mass cut at  $5000 M_{\odot}$ , which ensures that we will be complete in their recovery rate up to 200 Myr, but to look now at the number densities of objects per unit time ( $dN/dt$ ) as a function of increasing age. A decreasing rate of objects as a function of time  $dN/dt \propto t^{\delta}$  is historically interpreted as evidence for YSC dissolution within the galaxy (see Lamers 2009 for a short review). In Figure 19, top panel, we plot the change in number densities of class 1 and 2 (orange dots), class 3 (blue diamonds), and the whole sample (green triangles) using bins of the same width (left panel). The shadowed areas are the regions of the diagram where incompleteness mimics a rapid disappearance of objects. Therefore, these regions are excluded (e.g., see Lamers 2009). Overall, we observe that the disruption rate of class 3 systems is certainly more significant than class 1 and 2. When the two classes are analyzed together, the resulting disruption rate is coincident with the one of class 1 and 2, probably because this latter class is much more numerous.

##### 5.4.1. Cluster and Compact Association Evolution During the First 10 Myr

We also observe variations depending on the age range. In all the plots of Figure 19, we clearly see that during the first 10 Myr the number densities of objects in both class 1 and 2 and class 3 are significantly higher than in the range 10–100 Myr. The number densities per unit time declines roughly by factors of 4 and 3 for class 3 and class 1 and 2, respectively, when comparing the rate between the range 1 to



**Figure 19.** Number density of systems more massive than  $5000 M_{\odot}$  per unit time as a function of age using equally spaced temporal bins (bin size is 0.6 dex). The shaded areas show the regions of the diagrams that are affected by incompleteness and excluded from the analysis. The fit to each distribution within the age range 10–200 Myr is illustrated with a dashed line. The recovered slopes are included in the corresponding insets. The left panel illustrates the change in number density of the whole population (class 1, 2, and 3, green triangles), cluster candidates (class 1 and 2, orange dots), and compact associations (class 3, blue diamonds). The central panel shows the number density of clusters as a function of age within an inner and outer region. The two regions contain the same number of clusters with mass above  $5000 M_{\odot}$ . In the left panel, we split the sample into low mass ( $\log(M) \leq 3.9$ , magenta dots) and high mass ( $\log(M) > 3.9$ , green diamonds) clusters. See the text to follow the discussion of the results.

10 Myr and 10 to 100 Myr. Two different explanations are typically advocated to describe such a downtrend, either “infant mortality,” in the classic terminology introduced by Lada & Lada (2003), or contamination of our sample with systems that are not gravitationally bound (e.g., Gieles & Portegies Zwart 2011 on the difficulty of distinguishing bound from unbound young star-forming regions). In the first scenario, it is assumed that all stars form in bound clusters that rapidly dissolve because gas evacuated by stellar feedback has destabilized the gravitational potential of the system. In the second scenario, the rapid disruption is only caused by the limits of the method we are employing. Dynamical imprints of very young star-forming regions in the Milky Way (e.g., Wright et al. 2014) and Magellanic Clouds (Gouliermis et al. 2014) suggest that even massive OB associations, like Cygnus OB2, are not the result of an evolution from a gravitationally bound status but are formed already unbound. Stellar feedback, therefore, does not appear to be a major agent of cluster disruption (e.g., Longmore et al. 2014 for a review). In extragalactic studies, like the one performed on the LEGUS galaxies, we are not able to probe the boundness status of the objects we consider clusters. Objects older than 10 Myr and with effective radii of a few parsecs can be considered gravitationally bound since their crossing time is smaller than the age of the stars. However, younger objects, often still nested within the large star-forming regions where they have formed, are more difficult to classify. Because of the complexity involved in characterizing objects at these young ages, we restrict our  $dN/dt$  analysis to the age range 10–200 Myr.

#### 5.4.2. The Change in the Number Density of Clusters and Associations as a Function of Time, Mass, and Galactocentric Distance

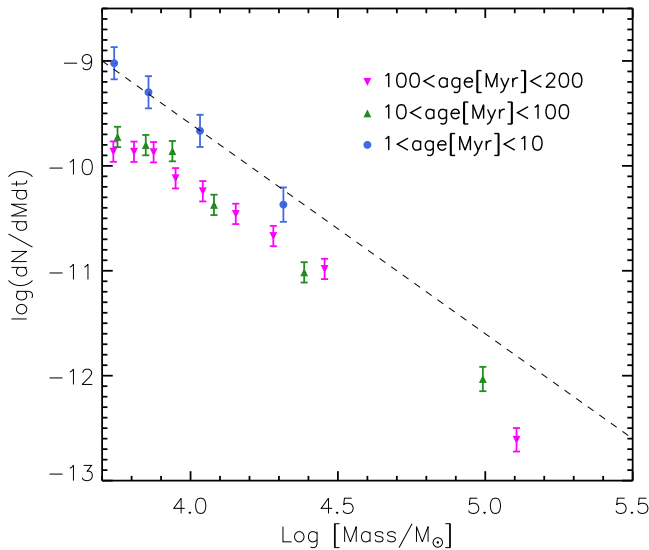
Between 10 and 200 Myr, the left plot of Figure 19 shows that the change in the number density of class 1 and 2 clusters is consistent with mild disruption. We do not observe any drastic dissolution of this class of objects up to 200 Myr. On the other hand, the number density of class 3 systems keeps declining by a factor of 3 in the same age range, suggesting, as already observed in the color–color diagrams analysis, their

rapid disappearance within 100 Myr (the slope we recover is  $\delta \sim -0.7$ ). These short timescales are in agreement with the clustering analysis performed by Grasha et al. (2015) on the class 3 population. The lack of clustering after 40 Myr may be caused by the quick dissolution of these systems. On the other hand, since class 1 and 2 survive longer, their lack of spatial clustering is possibly the result of the randomization of their positions because they move away from their birthplaces. The clustering analysis of stellar structures performed by Gouliermis et al. (2015) in another LEGUS galaxy, NGC 6503, finds that hierarchical clustered stellar structure disappear and distribute into the stellar field within 60 Myr. This suggests that class 3 objects are well nested within the hierarchical properties of star formation, while stellar clusters, even though have formed within the same hierarchically structured ISM, as gravitationally bound systems, may follow a different fate.

Recent theoretical and numerical works (e.g., Elmegreen & Hunter 2010; Kruijssen et al. 2011) point out that tidal forces exerted by GMC encounters in a hierarchical ISM can reproduce the steady decrease in cluster numbers over time as a result of a decreasing ISM density in each cluster’s environment as it drifts away from its birth place. Higher gas densities toward the centers of spiral galaxies or in starburst systems can increase the overall dissolution rate of YSCs.

In Figure 8, the color–color diagrams of the inner and outer cluster population of NGC 628 show interesting features, suggesting that the outer cluster population is preferentially older. We attempt here to investigate whether we observe any change in the recovered dissolution time as a function of galactocentric distances. We split our sample of class 1 and 2 objects into a central and outer bin containing the same number of objects. In the central panel of Figure 19, we show the recovered disruption rates for the inner and outer bins. We observe that the number density of clusters in the outer bin is consistent with being constant between 10 and 200 Myr ( $\delta \sim 0.0$ ), while in the inner bin we observe higher dissolution rates  $\delta \sim 0.3$ . The observed trends are consistent with theoretical expectations, suggesting a higher mass-independent disruption in denser galactic ISM, thus closer to the center of the spiral galaxy.





**Figure 20.** Number of clusters per mass and time unit ( $dN/dMdt$ ). Only clusters more massive than  $5000 M_\odot$  and classified as class 1 and 2 are included. Blue dots show the CMF for ages below 10 Myr. Green and magenta triangles show the CMF between 10 and 100 Myr and between 100 and 200 Myr, respectively. The dashed line is included to provide a reference for a power-law CMF with slope  $-2$ . See the text for details.

In the right panel of Figure 19, we probe the disruption rate of low (magenta dots) and high (green diamonds) mass cluster candidates. We observe that low mass clusters show evidence of mild disruption ( $\delta \sim -0.3$ ) while the number densities of clusters with mass similar to or larger than  $10^4 M_\odot$  are consistent with being constant ( $\delta \sim 0.0$ ). This result suggests the presence of a mass dependency in the disruption rate of clusters.

Another way to investigate cluster evolution as a function of the cluster mass is to encode in a bivariate distribution,  $g(M, t)$  the mass function and time dependence for formation and evolution of clusters (see Fall et al. 2009; Gieles 2009). The function  $g(M, t)$  is thus the number of clusters observed as a function of time and mass expressed as

$$g(M, t) = \frac{d^2N}{dMdt}. \quad (2)$$

Integrating  $g(M, t)$  over the mass provides the  $dN/dt$  distributions, while integrating  $g(M, t)$  over a time range provides the observed CMF.

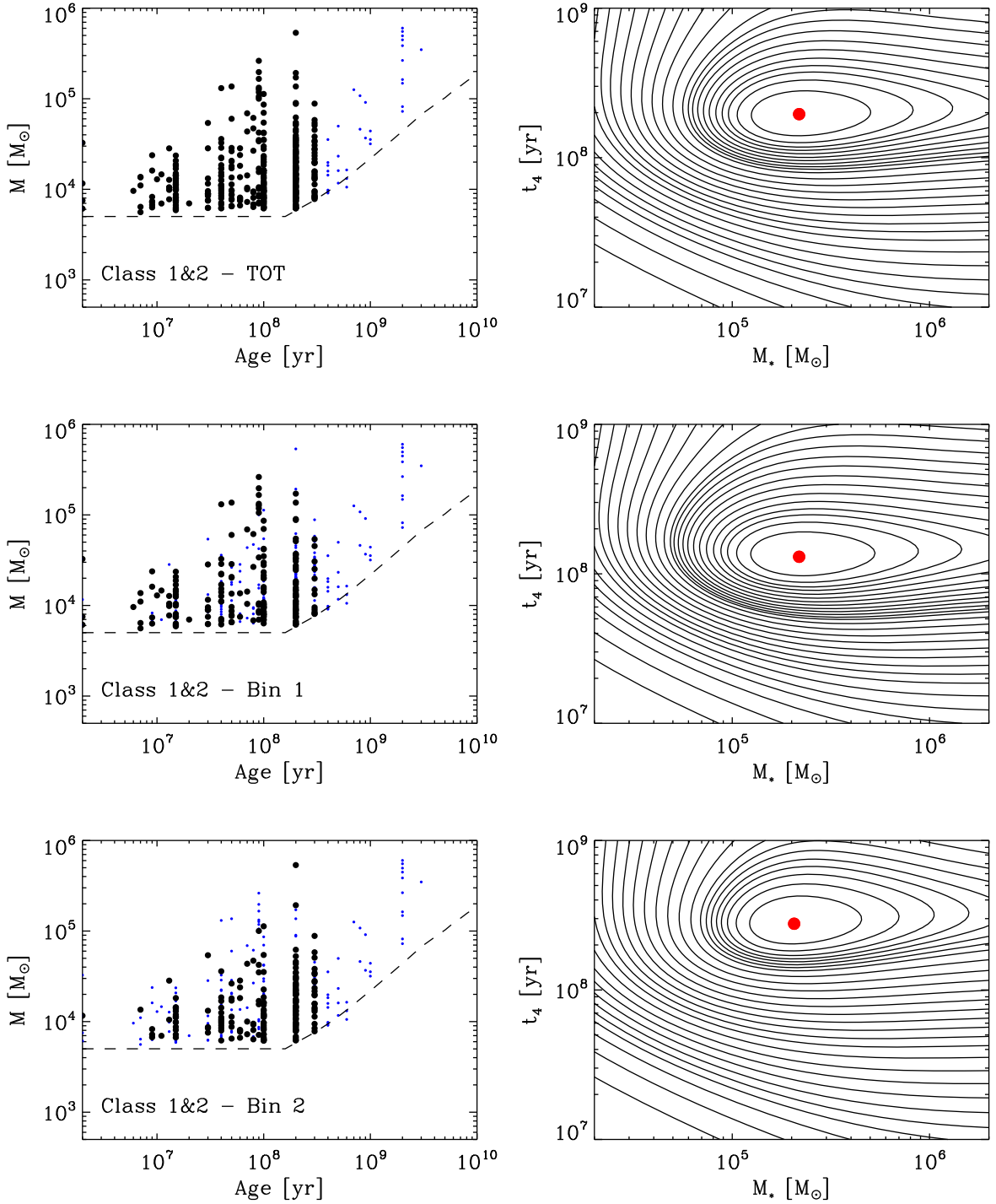
In Figure 20 we show the recovered mass distributions as a function of cluster age normalized by the corresponding age interval (i.e.,  $dN/dMdt$  diagnostic). The distribution is built to contain the same number of objects in each bin. The  $dN/dMdt$  diagnostic is sensitive to the evolution of the mass function with time. If the number of clusters rapidly declines independently of the cluster mass then the shape of the mass function will be unchanged but the distribution at each age interval will be shifted because the number of clusters is diminishing. On the other hand, if the cluster disruption timescale depends on cluster mass, then the aging CMFs should overlap at the mass ranges untouched by disruption and deviate where disruption of low mass clusters is significant. The plot in Figure 20 shows a clear offset of the CMF for clusters younger than 10 Myr. As already seen in the  $dN/dt$

distribution, the number of clusters at this age range is significantly higher. The  $dN/dMdt$  analysis provides further insights showing that the number of clusters younger than 10 Myr is higher at any mass range in units of time. The CMFs of the other two age ranges (10–100 and 100–200 Myr) overlap at masses larger than  $10^4 M_\odot$ , while deviations become significant at lower masses. These deviations suggest that the number of clusters with masses below  $10^4 M_\odot$  are decreasing, flattening the CMF. This trend is consistent with both the effect of detection limits causing the loss of an increasing number of low mass clusters at older ages and higher disruption rates of clusters below  $10^4 M_\odot$  (consistent with the trends observed in the right plot of Figure 19). As pointed out in Section 5.2, our mass cut of  $5000 M_\odot$  is very close to the completeness limits at 200 Myr (see Figure 14). Incompleteness can mimic a mass-dependent cluster dissolution, and we cannot exclude that the observed flattening is in part caused by incompleteness.

To understand whether the trends observed in the bivariate distribution of the CMF as a function of time are compatible with mass-dependent disruption we perform a maximum-likelihood fit to our data. We refer to Gieles (2009) for the formalism behind the fit and to Bastian et al. (2012) for an application to the YSC population in the M83 galaxy. The fit is performed assuming a Schechter function that describes the probability to form a cluster of mass  $M_i$  in the time interval  $t$  and  $t + dt$  as

$$\frac{d^2N}{dM_i dt} = A M_i^{-2} \exp\left(-\frac{M_i}{M_*}\right), \quad (3)$$

where  $M_i$  depends on time. If only cluster mass-dependent disruption is taken into account, the disruption time can be described as  $t_{\text{dis}} = t_0 M^\gamma$ , with  $t_0$  and  $\gamma$  depending on the galactic environment (see e.g., Lamers et al. 2005), and the mass evolution is described by  $M = (M_i^\gamma - \gamma t/t_0)^{1/\gamma}$  (Lamers et al. 2005; Gieles 2009). In the fit,  $\gamma$  is fixed to the average value found in local spirals ( $\gamma = 0.65$ ). The fitting algorithm thus finds the values of  $M_*$ ,  $t_0$ , and  $t_4$  that maximize the likelihood. The  $t_4$  is the timescale necessary to dissolve a cluster of  $M_4 = 10^4 M_\odot$ , and it depends on  $t_0$  as  $t_4 = (t_0/\gamma) M_4^\gamma$ . The results of maximum-likelihood fits performed on the YSC population of NGC 628 are shown in Figure 21. The fit is performed on class 1 and 2 clusters more massive than  $5000 M_\odot$ , brighter than the magnitude cut  $M_V = -6$  ( $V = 23.98$  mag), and younger than 300 Myr. The age–mass diagram on the left side panels shows the YSC population taken into account in the fit. The mass cut removes low mass objects up to an age of 200 Myr. The magnitude cut becomes important at older ages because it removes clusters more massive than  $5000 M_\odot$  at ages older than 200 Myr. The fit is done for the entire cluster population (top), and the inner (center) and outer (bottom) regions of the galaxy corresponding to two bins containing the same number of clusters. The maximum-likelihood fit performed on the entire cluster population finds  $M_* = 2.2 \times 10^5 M_\odot$  and a timescale for dissolution of a  $10^4 M_\odot$  cluster of  $t_4 = 190$  Myr. These values are in agreement with the  $M_*$  found in Section 5.2 and the evidence of slow disruption in the  $dN/dt$  analysis performed above. With the same technique, we derive  $M_*$  and  $t_4$  for clusters of the inner and outer regions containing the same number of objects, as defined above. The cluster population



**Figure 21.** Maximum-likelihood fit of the YSC population of NGC 628. Only class 1 and 2 clusters more massive than  $5000 M_{\odot}$  and younger than 300 Myr are included. A magnitude limit corresponding to the  $M_V = -6$  magnitude cut applied to select cluster candidates is also taken into account in the fitting. The left panels show the age–mass diagram of the sources included in the fit (black dots) overplotted on the whole class 1 and 2 sample (blue dots). The dashed line shows how the mass and magnitude limits select clusters at different age ranges. At young ages, the mass cut is limiting the sample (horizontal dashed line); at older ages the luminosity cut is more important (transversal dashed line at ages larger than 200 Myr). The right panels show the values of  $t_4$  and  $M_*$  that provide the maximum likelihood (red dots). In the top panels, the total YSC population of NGC 628 has been fitted. The central and bottom panels show the results of the fit performed on the inner and outer bin population as shown in Figure 18. See the text for details.

located within the inner bin has the same  $M_*$  value found for the entire galaxy ( $M_* = 2.2 \times 10^5 M_{\odot}$ ) and a timescale for dissolution of a  $10^4 M_{\odot}$  cluster of  $t_4 = 130$  Myr. The clusters in the outer region have a factor of 2 longer disruption timescales with  $t_4 = 270$  Myr and  $M_* = 2.0 \times 10^5 M_{\odot}$ . The maximum-

likelihood fitting analysis confirms the results and trends obtained from the analyses of the CMF and the  $dN/dt$  distributions. The average timescale for cluster disruption in the galaxy are long enough to produce a shallow  $dN/dt$  distribution. However, the age interval to which our analysis is sensitive too is long enough

that we should start to see the effect of disruption on low mass clusters. Indeed, in this respect the  $dN/dMdt$  analysis is a more sensitive diagnostic than the simple  $dN/dt$  one. The longer timescales for the disruption of clusters in the outer part of the galaxy are also in agreement with the expectation from theoretical works (e.g., Elmegreen & Hunter 2010; Kruijssen et al. 2011) and the results obtained for the M83 spiral galaxy (Bastian et al. 2012).

To verify the effect of incompleteness at the low mass end as a function of age, we repeat the maximum-likelihood fit using the same mass cut at  $5000 M_\odot$  together with a more conservative luminous cut, i.e.,  $M_v = 22.93$  mag, i.e., one magnitude brighter than previously done. The mass and magnitude cuts result in a selection of clusters more massive than  $5000 M_\odot$  up to 100 Myr and objects brighter than 22.93 mag at older ages. We obtain values of  $M_*$  and  $t^4$  within a factor of two from the analysis performed above with a less conservative magnitude cut. With a limiting brightness of  $M_v = -22.93$  mag, we retrieve  $M_* = 1.7 \times 10^5 M_\odot$  and  $t^4 = 170$  Myr for the entire sample, and  $M_* = 2.2 \times 10^5$  and  $1.4 \times 10^5 M_\odot$ , and  $t^4 = 100$  and 530 Myr for the central and out region of NGC 628, respectively. This further test reinforces the evidence, found above, of a mass-dependent component necessary to describe cluster disruption in NGC 628.

## 6. Discussion

### 6.1. The Shape of the CMF at the High Mass End and Constraints on the Cluster Formation Process

CLF and CMF are powerful tools to investigate the formation of YSCs.

In the spiral galaxy NGC 628, using the LEGUS data set, we recover the properties of the CLF from the UV to the NIR. The luminosity distributions of the cluster (class 1 and 2) population is close to the slope  $-2$ , while the compact stellar associations (class 3) appear to have steeper slopes. The analysis of increasing size of stellar aggregates (Elmegreen et al. 2006) also finds slopes close to  $-2$ . Hence, star formation is consistent with a hierarchical process, driven by turbulence, from the largest scales, i.e., star-forming complexes, down to the densest and smallest physical scales, i.e., star clusters. Our analysis also reveals some interesting variations in the CLF for both clusters and associations. At each waveband, we observe a steepening at the brightest end of the cumulative distributions. Extinction cannot explain the steepening at the brightest end of the CLF. If more luminous clusters are more extinguished than faint ones, it would imply preferential extinction as a function of luminosity, a trend not observed in studies of local galaxies.

The cause of the steepening of the CLF could be connected to the nature of the cluster formation process. The CLF is a direct observable tracer of the underlying CMF integrated over time. If the CMF is a pure power-law function of slope  $-2$  at any age and cluster dissolution is independent of the cluster mass, the resulting CLF should consistently be a function with the same slope. The steepening that we are observing at the bright end of the distributions suggests that we find fewer luminous (massive) clusters than expected for a pure power-law CMF.

Increasing evidence in the literature suggests that the YSC mass function is better described by a truncated power law. Johnson et al. (2017), studying the YSC population of M31, find that the CMF in this galaxy has a Schechter-type form,

with  $M_* \sim 1 \times 10^4 M_\odot$ . Cosmological simulations of the Milky Way type of galaxies, where YSCs are implemented as star formation units, show that the resulting YSC populations have CMF that are better described by a Schechter function and  $M_*$  scales as a function of SFR (Li et al. 2017).

In the case of NGC 628, both a steeper power law ( $\alpha \sim 2.1$ ) or a Schechter function ( $\alpha \sim 2.0$  and  $M_* = 2.0 \times 10^5 M_\odot$ ) can equally reproduce the observed CMF. We investigate whether the steepening observed in the mass function, which suggests a dearth of very massive systems (if a pure power-law function is assumed), is a sign of an underlying soft truncation. The difficulty in constraining the upper mass distribution of the CMF in local spirals has already been pointed out by Larsen (2006). The combination of cluster formation being a stochastic process (e.g., Adamo & Bastian 2015) and relatively low SFR combined with detection limits and rapidly fading luminosities above a few hundreds of Myr make it very challenging to put strong constraints on the shape of the cluster upper mass distribution, because of low number statistics.

### 6.2. Timescales for Cluster Disruption in NGC 628; Probing Cluster Evolution

Another fundamental aspect that gives important constraints on the formation and evolution of YSCs is understanding how they disrupt. Star-forming regions and stellar complexes are hierarchical in space and time (Efremov & Elmegreen 1998), thus their crossing times are comparable to their ages and appear to dissolve on timescales below 60 Myr (e.g., Pellerin et al. 2012; Crocker et al. 2015; Gouliermis et al. 2015). On the other hand, the fate of YSCs, forming in the densest peaks of these very regions, is not yet well understood.

In our analysis, we look at the number density of stellar systems between 1 and  $\sim 200$  Myr. We are not able to derive the dynamical status of the stars within our objects, but since the average size of our class 1 and 2 systems peaks at 3 pc (Ryon et al. 2017), their crossing times, for those older than 10 Myr, are much shorter than the stellar ages so we can consider them likely gravitationally bound. We see a clear decline between the number of objects during the first 10 Myr and the next age bin, suggesting a significant loss of both class 1 and 2 and class 3 systems. The recovered number densities of YSCs versus associations in the age range 10–200 Myr are significantly different. Compact stellar associations (class 3) rapidly decline in number and disappear on timescales ( $\sim 50$  Myr) comparable to those of hierarchically structured star-forming regions. Objects that we have identified as potentially bound clusters (class 1 and 2) show close to constant number densities ( $\delta \sim -0.2$ ) in the  $dN/dt$  analysis.

Small disruption rates within the first 100–200 Myr have also been reported in the literature for other local galaxies, e.g., in M31 (Fouesneau et al. 2014), LMC (Baumgardt et al. 2013), and M83 (Silva-Villa et al. 2014). However, we notice here that these results are not ubiquitous. A compilation of previous works (Adamo & Bastian 2015) clearly shows that disruption rates of YSCs may change significantly from galaxy to galaxy and become very high for YSCs in hostile environments like the Antennae system (Whitmore et al. 2010).

Our results suggest a steeper disruption slope for YSCs located in the inner portion of the galaxy ( $\delta \sim -0.3$ ). This finding is observationally supported by the difference in the number densities of YSCs in regions of the color-color diagrams (Figure 8) corresponding to young ages. In the inner



pointing, the number of YSCs is larger than in the outer pointing at young ages, while similar number densities are observed at older ages between the populations of the two pointings. If star formation has been constant in the last few hundreds of Myr, then these differences suggest a longer survival time in the outer regions. A larger coverage of the galaxy would certainly improve the results, as recently shown for M83 (e.g., Chandar et al. 2010, 2014; Bastian et al. 2012; Silva-Villa et al. 2014).

## 7. Conclusions

We present the methods and pipelines developed and applied to build uniform YSC catalogs of the LEGUS galaxies. Our method consists of a mixture of automated and visually optimized procedures, which take into account differences in the quality of the data, distance of the galaxy, coverage, and varying local background. We implement a quality-flag system, based on human and ML approaches under development, which describe the morphology of sources that are detected in at least four bands, and have luminosities in the visual band brighter than  $-6$  mag. We provide final YSC catalogs which include positions, CI of the source, photometry in the five LEGUS bands; ages, masses, extinctions, and uncertainties;  $\chi^2$  analysis including residuals; and visual classification flags. Cluster photometry is produced with the two most used methods in the literature, which consist of fixed aperture photometry corrected by (1) an average aperture correction derived from observed clusters in each band, or (2) a CI-based aperture correction as a function of wavelength. The SED of each source is then fitted with both deterministic and stochastic SSP models, which include a treatment for nebular emission. We used both Padova and Geneva stellar libraries and three different recipes for internal extinction.

We provide, as a proof of concept, a detailed description of all the steps necessary to produce the final cluster candidate catalogs, with parameters and assumptions optimized for the LEGUS target NGC 628. Two regions of the galaxy have been targeted by LEGUS, the inner and the outer one.

In the attempt to probe YSC formation and evolution, we analyze the luminosity and physical properties of the YSC population in NGC 628, using as reference catalog the one obtained with Padova evolutionary tracks, Kroupa IMF, solar metallicity, and Cardelli extinction law. A comparison between the cluster properties obtained with the two photometric methods (AV\_APCOR and CI\_BASED) shows some differences in the color-color distributions and recovered cluster physical properties. We conclude that significant deviations affect about 20% of the sources in common between the two catalogs. We perform the analysis for both photometric catalogs, but results are not affected by the choice of the catalog.

The color-color diagrams of class 1 and 2 systems (which, according to our visual classification scheme, are YSC candidates) and class 3 objects (likely compact stellar associations) show interesting differences in their density distributions. YSC candidates follow closely the SSP models at all ages, while stellar associations are more concentrated around the regions of the tracks younger than 50–60 Myr. We also compare the color properties of the YSCs and stellar associations in the inner and outer regions of NGC 628. We do not observe any significant difference between the number of stellar association distributed

along the evolutionary models, suggesting similar age ranges for the associations in these two pointings. On the other hand, in regions of the color-color diagram corresponding to young ages, the number density of YSCs changes between the inner and outer pointing. On average, the population in the outer field is older, probably reflecting a slower disruption rate in the outer part of the galaxy.

Thanks to the LEGUS multiband coverage, we produce a complete luminosity function analysis from the UV to the NIR. A power-law fit to the luminosity distributions yields slopes that are consistently close to  $-2$ . We also observe, for both YSC candidates and stellar associations, that the recovered slopes show a steepening from the shorter to the longer wavebands possibly consistent with wavelength-dependent fading effects. At the bright end of the CLF, at all wavelengths, we see a clear deviation from the power-law shape, suggesting a dearth of luminous clusters. The analysis of the CMF and Monte Carlo simulations of cluster populations suggest that the CMF can be described by a power-law function with a slope close to  $-2$  with a truncation at  $M_\star \sim 2 \times 10^5 M_\odot$ . However, due to the low number statistics, the solution of a pure power-law function with a slope of  $\alpha \sim -2.1$  cannot be discarded.

The analysis of the number densities of objects as a function of age shows different trends for YSC candidates and stellar associations. The numbers of stellar associations decline rapidly and they tend to disappear on short timescales as already observed in the color-color diagrams. On the other hand, bound systems do not show any drastic decline in their numbers between 10 and  $\sim 200$  Myr. We find evidence of a more significant cluster disruption rate in the inner region of NGC 628, in agreement with expectations of higher chances of encounters with GMCs and consistent with theoretical predictions of an environmental dependency in cluster disruption. We estimate that the timescale to disrupt a cluster with  $M = 10^4 M_\odot$  is shorter (130 Myr) closer to the center and significantly longer in the outer part of the galaxy (270 Myr). These timescales should be considered as lower limits, because we cannot exclude that incompleteness at older ages is affecting our results.

Since we do not observe significant disruption for cluster with masses above  $10^4 M_\odot$ , we conclude that the observed  $M_\star$  values are linked to the formation mechanism of the YSC population and not to their evolution.

The analysis performed on the YSC population of NGC 628 shows important evidence for our understanding of the formation and evolution of YSC and stellar associations within their host galaxy. Our morphologically based classification provides, for the first time, insights on the properties of our stellar systems within the framework of a hierarchically driven star formation process. YSCs and stellar associations form within star-forming regions and inherit the imprints of the turbulent status of the ISM as proved by their power-law slope  $-2$ . However, evidence suggests that at some physical scales, other physical parameters may play an important role and affect the shape of the CMF. After their formation, YSCs and associations seem to follow different evolution paths, with the former surviving untouched for a longer time frame, while associations disappear on timescales comparable to hierarchically organized star-forming regions.

These initial results show interesting trends, but only the sampling provided by a survey like LEGUS can allow us to address fundamental open questions related to YSC formation

and evolution. A large pool of different galactic environments will provide us with the means to investigate whether the CMF is universal, what physical properties affect the upper mass end of the mass function, what mechanism dominates cluster disruption, and whether cluster formation efficiency is a relevant quantity to describe cluster formation in the local universe.

We are thankful to the anonymous referee for comments and suggestions that have improved the manuscript. A.A. thanks Nate Bastian for sharing the codes used to make the density contours of the color–color diagrams and the maximum-likelihood analysis. Mark Gieles and Nate Bastian are thanked for comments on an early draft of the manuscript. A.A. acknowledges partial support from the Swedish Royal Academy. G.A. acknowledges support from the Science and Technology Facilities Council (ST/L00075X/1 and ST/M503472/1). C.D. acknowledges funding from the FP7 ERC starting grant LOCALSTAR (no. 280104). M.F. acknowledges support by the Science and Technology Facilities Council (grant number ST/L00075X/1). D.A.G. kindly acknowledges financial support by the German Research Foundation (DFG) through program GO 1659/3-2. A.H. thanks the Spanish MINECO for grant AYA2015-68012-C2-1. These observations are associated with program # 13364. Support for program # 13364 was provided by NASA through a grant from the Space Telescope Science Institute. Based on observations obtained with the NASA/ESA *Hubble Space Telescope*, at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

*Facility:* *HST* (WFC3, ACS).

## References

- Adamo, A., & Bastian, N. 2015, arXiv:1511.08212
- Adamo, A., Kruijssen, J. M. D., Bastian, N., Silva-Villa, E., & Ryon, J. 2015, *MNRAS*, **452**, 246
- Adamo, A., Östlin, G., Zackrisson, E., et al. 2010a, *MNRAS*, **407**, 870
- Adamo, A., Östlin, G., Zackrisson, E., et al. 2011, *MNRAS*, **415**, 2388
- Adamo, A., Smith, L. J., Gallagher, J. S., et al. 2012, *MNRAS*, **426**, 1185
- Adamo, A., Zackrisson, E., Östlin, G., & Hayes, M. 2010b, *ApJ*, **725**, 1620
- Annibali, F., Tosi, M., Aloisi, A., & van der Marel, R. P. 2011, *AJ*, **142**, 129
- Bastian, N., Adamo, A., Gieles, M., et al. 2011, *MNRAS*, **417**, L6
- Bastian, N., Adamo, A., Gieles, M., et al. 2012, *MNRAS*, **419**, 2606
- Baumgardt, H., Parmentier, G., Anders, P., & Grebel, E. K. 2013, *MNRAS*, **430**, 676
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bik, A., Lamers, H. J. G. L. M., Bastian, N., Panagia, N., & Romaniello, M. 2003, *A&A*, **397**, 473
- Bik, A., Östlin, G., Hayes, M., et al. 2015, *A&A*, **576**, L13
- Boutloukos, S. G., & Lamers, H. J. G. L. M. 2003, *MNRAS*, **338**, 717
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682
- Calzetti, D., Lee, J. C., Sabbi, E., et al. 2015, *AJ*, **149**, 51
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, **345**, 245
- Chandar, R., Fall, S. M., & Whitmore, B. C. 2015, *ApJ*, **810**, 1
- Chandar, R., Whitmore, B. C., Calzetti, D., & O’Connell, R. 2014, *ApJ*, **787**, 17
- Chandar, R., Whitmore, B. C., Kim, H., et al. 2010, *ApJ*, **719**, 966
- Crocker, A. F., Chandar, R., Calzetti, D., et al. 2015, *ApJ*, **808**, 76
- da Silva, R. L., Fumagalli, M., & Krumholz, M. 2012, *ApJ*, **745**, 145
- Efremov, Y. N., & Elmegreen, B. G. 1998, *MNRAS*, **299**, 588
- Ekström, S., Georgy, C., Eggenberger, P., et al. 2012, *A&A*, **537**, A146
- Eldridge, J. J., Izzard, R. G., & Tout, C. A. 2008, *MNRAS*, **384**, 1109
- Elmegreen, B. G., & Efremov, Y. N. 1997, *ApJ*, **480**, 235
- Elmegreen, B. G., Elmegreen, D. M., Chandar, R., Whitmore, B., & Regan, M. 2006, *ApJ*, **644**, 879
- Elmegreen, B. G., & Hunter, D. A. 2010, *ApJ*, **712**, 604
- Fall, S. M., Chandar, R., & Whitmore, B. C. 2005, *ApJL*, **631**, L133
- Fall, S. M., Chandar, R., & Whitmore, B. C. 2009, *ApJ*, **704**, 453
- Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, *RMxAA*, **49**, 137
- Fouesneau, M., Johnson, L. C., Weisz, D. R., et al. 2014, *ApJ*, **786**, 117
- Gieles, M. 2009, *MNRAS*, **394**, 2113
- Gieles, M. 2010, in ASP Conf. Ser. 423, *Galaxy Wars: Stellar Populations and Star Formation in Interacting Galaxies*, ed. B. Smith et al. (San Francisco, CA: ASP), 123
- Gieles, M., Bastian, N., Lamers, H. J. G. L. M., & Mout, J. N. 2005, *A&A*, **441**, 949
- Gieles, M., Larsen, S. S., Bastian, N., & Stein, I. T. 2006, *A&A*, **450**, 129
- Gieles, M., & Portegies Zwart, S. F. 2011, *MNRAS*, **410**, L6
- Glatt, K., Grebel, E. K., & Koch, A. 2010, *A&A*, **517**, A50
- Goddard, Q. E., Bastian, N., & Kennicutt, R. C. 2010, *MNRAS*, **405**, 857
- Gouliermis, D. A., Hony, S., & Klessen, R. S. 2014, *MNRAS*, **439**, 3775
- Gouliermis, D. A., Thilker, D., Elmegreen, B. G., et al. 2015, *MNRAS*, **452**, 3508
- Grasha, K., Calzetti, D., Adamo, A., et al. 2015, *ApJ*, **815**, 93
- Haas, M. R., Gieles, M., Scheepmaker, R. A., Larsen, S. S., & Lamers, H. J. G. L. M. 2008, *A&A*, **487**, 937
- Holtzman, J. A., Faber, S. M., Shaya, E. J., et al. 1992, *AJ*, **103**, 691
- Hunter, D. A., Elmegreen, B. G., Dupuy, T. J., & Mortonson, M. 2003, *AJ*, **126**, 1836
- Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2015, *ApJ*, **802**, 127
- Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2017, *ApJ*, **839**, 78
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, **50**, 531
- Kennicutt, R. C., Jr., Edgar, B. K., & Hodge, P. W. 1989, *ApJ*, **337**, 761
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Kruijssen, J. M. D. 2012, *MNRAS*, **426**, 3008
- Kruijssen, J. M. D. 2015, *MNRAS*, **454**, 1658
- Kruijssen, J. M. D., Pelupessy, F. I., Lamers, H. J. G. L. M., Portegies Zwart, S. F., & Icke, V. 2011, *MNRAS*, **414**, 1339
- Krumholz, M. R., Adamo, A., Fumagalli, M., et al. 2015, *ApJ*, **812**, 147
- Krumholz, M. R., Fumagalli, M., da Silva, R. L., Rendahl, T., & Parra, J. 2015, *MNRAS*, **452**, 1447
- Lada, C. J., & Lada, E. A. 2003, *ARA&A*, **41**, 57
- Lamers, H. J. G. L. M. 2009, *Ap&SS*, **324**, 183
- Lamers, H. J. G. L. M., Gieles, M., Bastian, N., et al. 2005, *A&A*, **441**, 117
- Larsen, S. S. 1999, *A&AS*, **139**, 393
- Larsen, S. S. 2002, *AJ*, **124**, 1393
- Larsen, S. S. 2006, arXiv:astro-ph/0606625
- Larsen, S. S. 2009, *A&A*, **494**, 539
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, **123**, 3
- Lelièvre, M., & Roy, J.-R. 2000, *AJ*, **120**, 1306
- Li, H., Gnedin, O. Y., Gnedin, N. Y., et al. 2017, *ApJ*, **834**, 69
- Longmore, S. N., Kruijssen, J. M. D., Bastian, N., et al. 2014, in *Protostars and Planets VI*, ed. H. Beuther et al. (Tucson, AZ: Univ. of Arizona Press), 291
- Mafz Apellániz, J., & Úbeda, L. 2005, *ApJ*, **629**, 873
- Mistani, P. A., Sales, L. V., Pillepich, A., et al. 2016, *MNRAS*, **455**, 2323
- Mora, M. D., Larsen, S. S., Kissler-Patig, M., Brodie, J. P., & Richtler, T. 2009, *A&A*, **501**, 949
- Pellerin, A., Meyer, M. M., Calzetti, D., & Harris, J. 2012, *AJ*, **144**, 182
- Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, *ARA&A*, **48**, 431
- Randriamanakoto, Z., Väisänen, P., Ryder, S., et al. 2013, *MNRAS*, **431**, 554
- Reines, A. E., Nidever, D. L., Whelan, D. G., & Johnson, K. E. 2010, *ApJ*, **708**, 26
- Rosolowsky, E. 2005, *PASP*, **117**, 1403
- Ryon, J. E., Bastian, N., Adamo, A., et al. 2015, *MNRAS*, **452**, 525
- Ryon, J. E., Gallagher, J. S., Smith, L. J., et al. 2017, *ApJ*, in press (arXiv:1705.02692)
- San Roman, I., Sarajedini, A., Holtzman, J. A., & Garnett, D. R. 2012, *MNRAS*, **426**, 2427
- Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, **737**, 103
- Silva-Villa, E., Adamo, A., Bastian, N., Fouesneau, M., & Zackrisson, E. 2014, *MNRAS*, **440**, L116
- Swinbank, A. M., Smail, I., Longmore, S., et al. 2010, *Natur*, **464**, 733
- Tang, J., Bressan, A., Rosenfield, P., et al. 2014, *MNRAS*, **445**, 4287
- Thilker, D. A., Walterbos, R. A. M., Braun, R., & Hoopes, C. G. 2002, *AJ*, **124**, 3118
- Vázquez, G. A., & Leitherer, C. 2005, *ApJ*, **621**, 695
- Whitmore, B. C., Chandar, R., Bowers, A. S., et al. 2014, *AJ*, **147**, 78
- Whitmore, B. C., Chandar, R., Schweizer, F., et al. 2010, *AJ*, **140**, 75

- Whitmore, B. C., Schweizer, F., Kundu, A., & Miller, B. W. 2002, [AJ](#), **124**, 147
- Whitmore, B. C., Schweizer, F., Leitherer, C., Borne, K., & Robert, C. 1993, [AJ](#), **106**, 1354
- Whitmore, B. C., Zhang, Q., Leitherer, C., et al. 1999, [AJ](#), **118**, 1551
- Wofford, A., Charlot, S., Bruzual, G., et al. 2016, [MNRAS](#), **457**, 4296
- Wright, N. J., Parker, R. J., Goodwin, S. P., & Drake, J. J. 2014, [MNRAS](#), **438**, 639
- Zackrisson, E., Bergvall, N., Olofsson, K., & Siebert, A. 2001, [A&A](#), **375**, 814
- Zackrisson, E., Rydberg, C.-E., Schaerer, D., Östlin, G., & Tuli, M. 2011, [ApJ](#), **740**, 13